CUK-H02334-59-P01490t

AMERICAN
JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

59

R. BRAUER
UNIVERSITY OF TORONTO

H. WEYL
THE INSTITUTE FOR ADVANCED STUDY

F. D. MURNAGHAN
JOHNS HOPKINS UNIVERSITY

H. WHITNEY
HARVARD UNIVERSITY

A. WINTNER
THE JOHNS HOPKINS UNIVERSITY

510.5

WITH THE COÖPERATION OF

V. G. GROVE
M. H. HEINS
D. C. LEWIS
T. RADÓ
H. S. WALL
R. BAER
J. DOUGLAS
W. HUREWICZ
N. LEVINSON
G. PALL
S. BOCHNER
D. MONTGOMERY
C. L. SIEGEL
D. C. SPENCER
M. A. ZORN

PUBLISHED UNDER THE JOINT AUSPICES OF
THE JOHNS HOPKINS UNIVERSITY
AND
AMERICAN MATHEMATICAL SOCIETY

VOLUME LXVII
1945

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

9/4/28

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

IN[illegible]X

PAGE

AGNEW, RALPH PALMER. Spans in Lebesgue and uniform spaces of translations of peak functions, ... 431
BAER, REINHOLD. The homomorphism theorems for loops. 450
BARKER, CHARLES B., JR. The Lagrange multiplier rule for two dependent and two independent variables, ... 256
BEGLE, EDWARD G. Duality theorems for generalized manifolds, . 59
BERNSTEIN, B. A. See Foster and Bernstein, page 329.
BRAUER, RICHARD. On the representation of a group of order g in the field of the g-th roots of unity, ... 461
BRUNK, H. D. Dirichlet series meromorphic in a half-plane, . . . 486
CHEN, KIEN-KWONG. Some relations between the behavior of a function and the absolute summability of its Fourier series, ... 239
CHEN, KIEN-KWONG. A generalization of Hardy's theorem with an application to the absolute summability of Fourier series, . . 249
CHEN, KIEN-KWONG. Criteria for the absolute convergence of a Fourier series at a given point, ... 285
DE CICCO, JOHN. See Kasner and De Cicco, page 157.
DUFFIN, R. J. and SCHAEFFER, A. C. Power series with bounded coefficients, ... 141
FIALKOW, AARON. Conformal classes of surfaces, ... 583
FOSTER, ALFRED L. and BERNSTEIN, B. A. A dual-symmetric definition of field, ... 329
GAMA, LELIO I. Notion de proximité et espaces à structure sphéroïdale, 42
GRIFFITHS, L. W. Universal functions of polygonal numbers, III, . 443
HAVILAND, E. K. On the asymptotic behavior of the Riemann ξ-function, 411
HEINS, MAURICE H. The problem of Milloux for functions analytic throughout the interior of the unit circle, ... 212
HESTENES, MAGNUS R. A theory of critical points, ... 521
HILL, J. D. Nörlund methods of summability that include the Cesàro methods of all positive orders, ... 94
HUA, L. K. and SHIH, W. T. On the lack of an Euclidean algorithm in $R(\sqrt{61})$, ... 209
JACKSON, S. B. The four-vertex theorem for surfaces of constant curvature, ... 563
JAC[illegible]SON, N. The radical and semi-simplicity for arbitrary rings, . 300
KASNER, EDWARD and DE CICCO, JOHN. Geometry of scale curves in conformal maps, ... 157

PAGE

Ladue, Mary Elizabeth. Note concerning the conformal and equilong geometries of fourth and fifth order Horn sets, ... 155

Lee, Hwa-Chung. On even-dimensional skew-metric spaces and their groups of transformations, ... 321

Levitzki, Jakob. Solution of a problem of G. Koethe, . . . 437

Livingood, John. A partition function with the prime modulus $P < 3$, 194

Marden, Morris. Axisymmetric harmonic vectors, ... 109

Martin, Monroe H. The rectilinear motion of a gas II, . . . 407

Pólya, G. and Szegö, G. Inequalities for the capacity of a condenser, 1

Reiner, M. A mathematical theory of dilatancy, ... 350

Sathe, L. G. On a congruence property of the divisor function, . . 397

Schaeffer, A. C. See Duffin and Schaeffer, page 141.

Schur, Issai. On Faber polynomials, ... 33

Schur, Issai. Ein Satz ueber quadratische Formen mit komplexen Koeffizienten, ... 472

Shanks, M. E. Monotone decompositions of continua, ... 99

Sheffer, I. M. An extension of a Perron system of linear equations in infinitely many unknowns, ... 123

Shih, W. T. See Hua and Shih, page 209.

Szász, Otto. On Lebesgue summability and its generalization to integrals, ... 389

Szegö, G. See Pólya and Szegö, page 1.

Tschen, Y. Why. Algebraisation of plane absolute geometry, . . 363

Valentine, F. A. A Lipschitz condition preserving extension for a vector function, ... 83

Whyburn, G. T. Extensions of plane continua mappings, . . . 505

Wilkins, J. Ernest, Jr. The contact of a cubic surface with a ruled surface, ... 71

Wintner, Aurel. Gibbs' phenomenon and the prime number theorem, 167

Wintner, Aurel. Number-theoretical almost-periodicities, . . 173

Wintner, Aurel. The densities of ideal classes and the existence of unities in algebraic number fields, ... 235

Wintner, Aurel. The non-local existence problem of ordinary differential equations, ... 277

Wintner, Aurel. Small perturbations, ... 417

Wintner, Aurel. Mean-values of arithmetical representations, . . 481

Both kinds of symmetrization increase (or leave unchanged) r_P, *preserve* A *and diminish (or leave unchanged)* L, $\bar{r}$, C, *and* Λ.

This statement is very much condensed. Let us consider first the symmetrization with respect to a line. Then the statement regarding A is trivial and that regarding L is classical; all the other statements are new to our knowledge. (In considering r_P we have to symmetrize not only s but also an infinitesimal circle including P. Thus, if we choose the line of symmetrization through P, P is not moved at all. There is an analogous remark about the next case, concerning the point of symmetrization.)

In case of symmetrization with respect to a point the statement about A is trivial and that concerning L is equivalent to the classical isoperimetric theorem. The statements about r_P and $\bar{r}$ are equivalent to the area theorems of Bieberbach [see Pólya and Szegö **14**, vol. 2, p. 21, problem 125, p. 22, problem 126]. The part concerned with Λ was suggested by Lord Rayleigh [**16**; **17**, p. 345] and proved by Faber [**5**]. The proof of the last author was rediscovered by E. Krahn [**7**, **8**] and preceded by a paper of R. Courant [**4**] who considered a kind of symmetrization different from that of Steiner. Finally the part concerning C is new so far as we know.

The following inequalities are equivalent to some of these statements (except the last one which is of a different nature, cf. for instance Pólya and Szegö **14**, vol. 2, p. 21, problem 124):

$$(1.2) \qquad r_P \leqq (A/\pi)^{1/2} \leqq (\pi/2)C, \qquad (A/\pi)^{1/2} \leqq \bar{r} \leqq L/(2\pi).$$

These inequalities and various special cases suggest that

$$(1.3) \qquad (\pi/2)C \leqq \bar{r}$$

which however we were not able to prove. We have shown in a former investigation [**15**] that $C \leqq \bar{r}$.

5. Returning again to the three-dimensional case we state the following further relation of the capacity C of a closed surface σ_0 to certain geometrical properties of σ_0 [J. C. Maxwell **9**, p. 117, Szegö **21**]:

Consider all closed surfaces σ_0 *with given diameter. The capacity* C *of* σ_0 *is a maximum if and only if* σ_0 *is a sphere.*

By the diameter of σ_0 we mean the greatest distance between any two of its points. Since the capacity is an increasing set-function it is sufficient to prove this theorem for *convex* surfaces. While the result of Poincaré-Faber-

Szegö gives a lower bound for C the theorem mentioned furnishes an upper bound for the same quantity.

In another part of the present paper we prove the following refinement of this theorem:

Let the inner surface σ_0 *of a condenser of capacity* C *be convex and the outer surface* σ_1 *be parallel to* σ_0. *Let* h *be the distance between* σ_0 *and* σ_1, A *the area of the surface* σ_0 *and* M *the integral of the mean curvature of* σ_0. *Then*

$$A/(4\pi h) < C < A/(4\pi h) + M/(4\pi). \tag{1.4}$$

For concentric spheres the second inequality becomes an equality.

It is customary to regard the left-hand side of (1.4) as an approximation to C for small h; see for instance Maxwell **9**, p. 140. On the other hand R. Clausius [**3**, p. 43] suggested an approximate calculation of the electric density according to which the right-hand side of (1.4) would furnish *a better approximation than the left-hand side.*

Concerning the limiting case $h \to \infty$ see **21**, p. 425.

We also prove the analogue of the inequalities (1.4) for the logarithmic potential.

6. The remaining part of this paper is devoted to the study of the capacity in interesting special cases, especially in that of an *ellipsoid.* Let us consider first the capacity C (see **1.2** above) of any convex solid S. Let V denote the volume of S, A the area of the surface of S, and M the integral of the mean curvature extended over the surface of S. It is well known [**2**, p. 110] that

$$(3V/(4\pi))^{1/3} < (A/(4\pi))^{1/2} < M/(4\pi) \tag{1.5}$$

unless S is a sphere. As a limiting case of the inequalities which we discussed in **1.3** and **1.5** we have

$$(3V/(4\pi))^{1/3} < C < M/(4\pi) \tag{1.6}$$

unless S is a sphere. For the special case of an ellipsoid with semi-axes a, b, c we prove the following refinement of (1.6):

(1.7)

$$(3V/(4\pi))^{1/3} < \tfrac{1}{3}\{(bc)^{1/2} + (ca)^{1/2} + (ab)^{1/2}\} < C < \tfrac{1}{3}(a + b + c) < M/(4\pi)$$

unless $a = b = c$, that is, unless the ellipsoid is a sphere. The proofs are based on some representations of the capacity of an ellipsoid, related to a representation of its area, which played a role in a recent investigation of the first author [**11**].

7. The quantity $(A/(4\pi))^{1/2} = R_A$ occurring in (1.5) is obviously the radius of the sphere whose surface is equal to that of the given solid. In connection with the inequalities (1.5) and (1.6) we observe that the capacity C is "not comparable" with R_A, that is, for certain solids C is greater and for others smaller than R_A. A. Russell [**19**], following a hint of Lord Kelvin, gives a numerical discussion of the ratio of these quantities in various interesting cases especially for ellipsoids of revolution (spheroids). Let β be the numerical excentricity of the revolving ellipse; then Russell proves that the expansion of the ratio C/R_A in a power series in β begins thus:

$$C/R_A = 1 \pm (2/945)\beta^6 + \cdots \tag{1.8}$$

where the signs ± hold according to the case of a prolate and an oblate ellipsoid, respectively. [Cf. also the earlier papers of K. Aichi **1** and Lord Rayleigh **18.**]

We prove that for an ellipsoid of revolution

$$C/R_A \gtrless 1 \tag{1.9}$$

according as it is prolate or oblate. More precisely, we prove that the expansion of the quantity $(R_A/C)^2$ in a power series in β^2 has only negative and positive coefficients, respectively, except for the constant term 1 in the first case and the two subsequent coefficients which are zero in both cases.

8. The capacity of an ellipsoid is expressed by an elliptic integral [(6.4)]. In particular the capacity of an elliptic disk is the arithmetic-geometric mean of the semi-axes a, b multiplied by $2/\pi$. Hence the two elliptic disks with semi-axes a, b and $(a+b)/2$, $(ab)^{1/2}$, respectively, have the same capacity. Using Landen's transformation we find a somewhat analogous transition from one ellipsoid to another preserving the capacity.

9. In Section **2** we prove that the capacity of a condenser is diminished by applying the Steiner symmetrization; in Section **3** we show the same for the Schwarz symmetrization. Section **4** is concerned with the two-dimensional case. In Section **5** we give the proof of the inequalities (1.4). Sections **6-9** deal with the special case of an ellipsoid, Section **10** with some further special cases.

For the sake of simplicity we assume that all surfaces and lines considered below are composed of a finite number of analytic pieces. It does not seem to be doubtful that this condition can be replaced by some more general one without altering the main line of our argument. Also our argument does not change if the solid bounded by σ_0 is multiply connected or consists of several separated pieces. This deserves to be emphasized since such topological differences may require essential modifications of the proof; see for instance Pólya **12**.

2. Steiner symmetrization. In this section we prove that the capacity of a condenser is diminished by the Steiner symmetrization.

1. The main idea of the proof is very simple; it can be explained by using the notation of **1. 2.** We denote by σ'_0 and σ'_1 the surfaces arising from σ_0 and σ_1 by symmetrization. Let $\psi = u = u(x, y, z)$ be the (harmonic) function for which the minimum C of the integral (1. 1) is attained. We wish to construct a function $\psi' = v = v(x, y, z)$ defined in the field between σ'_0 and σ'_1, equal to 0 on σ'_0 and to 1 on σ'_1, and such that

$$W(u) \geqq W'(v). \tag{2.1}$$

Here W' denotes the expression analogous to (1. 1) for the symmetrized condenser; the integration in W' is extended over the field between σ'_0 and σ'_1.

If such a construction is possible the assertion is proved. Indeed the left-hand side of (2. 1) is equal to C while the right-hand side is not less than the capacity C' of the symmetrized condenser.

As to the construction of v we consider the level surfaces σ_λ of u defined by the equation

$$u = \lambda; \qquad 0 \leqq \lambda \leqq 1. \tag{2.2}$$

Since u is harmonic we have $0 \leqq u \leqq 1$ in the field between σ_0 and σ_1. Thus the level surfaces (2. 2) are all in this field and exhaust it completely; that is, through each point of the field passes a definite surface (2. 2). Observe that σ_β contains σ_α in its interior if $\beta > \alpha$. Let σ'_λ be the surface arising from σ_λ by the symmetrization in question. Through each point of the field between σ'_0 and σ'_1 passes a definite surface σ'_λ; an easy geometric consideration confirms that σ'_β contains σ'_α if σ_β contains σ_α. Now we define $\psi' = v$ by "transplanting the parameter," that is, we ascribe to $\psi' = v$ the value λ at each point of the surface σ'_λ.

The remaining part of the proof is devoted to the verification of (2. 1).

2. We assume that the plane ϵ with respect to which we symmetrize coincides with the x, y-plane. Let

$$z = z_\mu(x, y, \lambda), \qquad (\mu = 1, 2, \cdots, 2m), \tag{2.3}$$

be the portions of the surface σ_λ in the neighborhood of the points at which the vertical line through $(x, y, 0)$ intersects σ_λ and suppose that $z_1 > z_2 > z_3 > \cdots > z_{2m}$. The number $2m = 2m(x, y)$ of these points is finite. We prove that the total contribution of these $2m$ points to the Dirichlet integral $W(u)$ is not less than the contribution of the corresponding two points of the symmetrized surface σ'_λ to the integral $W'(v)$.

Denoting by γ_μ the angle between the exterior normal vector of σ_λ at (x, y, z_μ) and the positive z-axis we have

$$\cos \gamma_\mu = \pm \{1 + (\partial z_\mu/\partial x)^2 + (\partial z_\mu/\partial y)^2\}^{-1/2} \tag{2.4}$$

with the + or — sign according as μ is odd or even. Now grad u is parallel to the exterior normal vector of σ_λ and $|\operatorname{grad} u| = \partial\lambda/\partial n$ so that

$$(\partial\lambda/\partial n) \cos \gamma_\mu = \partial\lambda/\partial z_\mu. \tag{2.5}$$

Here $(-1)^{\mu-1}\partial\lambda/\partial z_\mu > 0$. Hence the contribution of the μ-th point to $W(u)$ will be

$$\begin{aligned}(\partial\lambda/\partial n)^2 dx\, dy\, dz &= (\partial\lambda/\partial n)^2 dx\, dy\, |dz_\mu| \\ &= (\cos \gamma_\mu)^{-2}(\partial\lambda/\partial z_\mu)^2 dx\, dy\, |\partial z_\mu/\partial\lambda|\, d\lambda \\ &= \frac{1 + (\partial z_\mu/\partial x)^2 + (\partial z_\mu/\partial y)^2}{|\partial z_\mu/\partial\lambda|}\, dx\, dy . d\lambda,\end{aligned} \tag{2.6}$$

where $d\lambda > 0$, from which we conclude that the total contribution is

$$dx\, dy\, d\lambda \sum_{\mu=1}^{2m} \frac{1 + (\partial z_\mu/\partial x)^2 + (\partial z_\mu/\partial y)^2}{|\partial z_\mu/\partial\lambda|}. \tag{2.7}$$

The corresponding quantity for the symmetrized domain can be obtained from this result by substituting $m = 1$ and replacing z_1 and $-z_2$ by

$$\tfrac{1}{2}(z_1 - z_2 + z_3 - z_4 + \cdots + z_{2m-1} - z_{2m}). \tag{2.8}$$

We find

$$dx\, dy\, d\lambda \frac{4 + (\sum_{\mu=1}^{2m}(-1)^{\mu-1}\partial z_\mu/\partial x)^2 + (\sum_{\mu=1}^{2m}(-1)^{\mu-1}\partial z_\mu/\partial y)^2}{\sum_{\mu=1}^{2m}(-1)^{\mu-1}\partial z_\mu/\partial\lambda}. \tag{2.9}$$

Our assertion follows from the elementary inequality

$$(2.10)\qquad \sum_{\mu=1}^{M} \frac{1+p_\mu^2+q_\mu^2}{r_\mu} \geqq \frac{4+(\sum_{\mu=1}^{M} p_\mu)^2+(\sum_{\mu=1}^{M} q_\mu)^2}{\sum_{\mu=1}^{M} r_\mu}$$

where p_μ, q_μ, r_μ are arbitrary real numbers, $r_\mu > 0$, $M \geqq 2$. This inequality is a simple consequence of the inequalities of Cauchy and Minkowski since

$$(2.11)\qquad \sum_{\mu=1}^{M} \frac{1+p_\mu^2+q_\mu^2}{r_\mu} \sum_{\mu=1}^{M} r_\mu \geqq \{\sum_{\mu=1}^{M} (1+p_\mu^2+q_\mu^2)^{1/2}\}^2$$

and

$$(2.12)\qquad \sum_{\mu=1}^{M} (1+p_\mu^2+q_\mu^2)^{1/2} \geqq \{(\sum_{\mu=1}^{M} 1)^2+(\sum_{\mu=1}^{M} p_\mu)^2+(\sum_{\mu=1}^{M} q_\mu)^2\}^{1/2}.$$

In our case $M = 2m \geqq 2$ so that $(\sum_{\mu=1}^{M} 1)^2 \geqq 4$. There is equality in (2.10) only for $M = 2$, $p_1 = p_2$, $q_1 = q_2$, $r_1 = r_2$.

3. Schwarz symmetrization. We now prove that the capacity of a condenser is diminished (or remains unchanged) by applying the process of the Schwarz symmetrization. Let the x-axis be the axis of the symmetrization. We retain the notations u, σ_λ introduced in Section 2 and define σ'_λ, v in the same way as in Section 2 replacing only the Steiner symmetrization by the Schwarz symmetrization.

1. We denote the area of the cross-section of σ_λ corresponding to a certain value of x by $A = A(x, \lambda)$. For the sake of simplicity we assume that this cross-section is bounded by a single closed curve which will be denoted by $s = s(x, \lambda)$.

We first compare $s(x, \lambda)$ and $s(x + dx, \lambda)$. Let $(\eta u_y, \eta u_z)$ be the normal vector of $s(x, \lambda)$ at (x, y, z) the end-point of which is on the projection of $s(x + dx, \lambda)$ onto the plane of $s(x, \lambda)$. Then

$$u(x + dx, y + \eta u_y, z + \eta u_z) = \lambda,$$

$$(3.1)\qquad u_x dx + \eta(u_y^2 + u_z^2) = 0, \qquad \eta = -u_x(u_y^2 + u_z^2)^{-1} dx,$$

so that

$$(3.2)\qquad |A_x| \leqq \int_{s(x,\lambda)} |u_x| (u_y^2 + u_z^2)^{-1/2} ds$$

where ds denotes the arc element of $s(x, \lambda)$.

On the other hand let $(\eta' u_y, \eta' u_z)$ be the normal vector of $s(x, \lambda)$ at (x, y, z) with end-point on $s(x, \lambda + d\lambda)$, $d\lambda > 0$. Then

$$(3.3) \qquad u(x, y + \eta' u_y, z + \eta' u_z) = \lambda + d\lambda, \qquad \eta'(u_y^2 + u_z^2) = d\lambda$$

so that

$$(3.4) \qquad |A_\lambda| = \int_{s(x,\lambda)} (u_y^2 + u_z^2)^{-1/2} ds.$$

2. Let us consider the ring-shaped domain between σ_λ and $\sigma_{\lambda+d\lambda}$, $d\lambda > 0$, which is cut off by the two symmetrizing planes through $(x, 0, 0)$ and $(x + dx, 0, 0)$. We prove as in Section **2** that the contribution of this domain to $W(u)$ is not less than the contribution of the corresponding part of the symmetrized domain to $W'(v)$.

Let (x, y, z) be an arbitrary point of σ_λ, that is $u(x, y, z) = \lambda$. Then

$$(3.5) \qquad \partial z/\partial x = -u_x/u_z, \qquad \partial z/\partial y = -u_y/u_z, \qquad \partial z/\partial \lambda = 1/u_z$$

and so from (2.6) we obtain the contribution

$$(3.6) \qquad dx\, d\lambda \int \frac{1 + u_x^2/u_z^2 + u_y^2/u_z^2}{1/|u_z|} |dy|$$

the integration being extended over $s(x, \lambda)$. But if x and λ are fixed, $u_y dy + u_z dz = 0$; hence

$$(3.7) \qquad |dy| = |u_z| (u_y^2 + u_z^2)^{-1/2} ds$$

and (3.6) becomes

$$(3.8) \qquad dx d\lambda \int_{s(x,\lambda)} (u_x^2 + u_y^2 + u_z^2)(u_y^2 + u_z^2)^{-1/2} ds.$$

3. The geometric condition defining the symmetrized surface σ'_λ can be expressed in the form $A(x, \lambda) = \pi(y^2 + z^2)$ and this equation defines the symmetrized function $\lambda = v(x, y, z)$. The intersection of σ'_λ with the plane through $(x, 0, 0)$ perpendicular to the x-axis is a circle with area A. We find easily that

$$(3.9) \qquad v_x = -A_x/A_\lambda, \qquad v_y = 2\pi y/A_\lambda, \qquad v_z = 2\pi z/A_\lambda$$

and so the corresponding contribution to (3.8) will be

$$\begin{aligned}(3.10) \qquad & dx d\lambda \{A_x^2 A_\lambda^{-2} + 4\pi^2(y^2 + z^2) A_\lambda^{-2}\}\{4\pi^2(y^2 + z^2) A_\lambda^{-2}\}^{-1/2} 2\pi (A/\pi)^{1/2} \\ & = dx d\lambda (A_x^2 + 4\pi A) A_\lambda^{-1}.\end{aligned}$$

Thus we have to prove that

$$(3.11)\qquad (A_x^2 + 4\pi A)A_\lambda^{-1} \leqq \int_{s(x,\lambda)} (u_x^2 + u_y^2 + u_z^2)(u_y^2 + u_z^2)^{-1/2} ds.$$

We use (3.2) and (3.4). By Schwarz's inequality

$$(3.12)\qquad A_x^2 \leqq A_\lambda \int_{s(x,\lambda)} u_x^2 (u_y^2 + u_z^2)^{-1/2} ds,$$

and by the isoperimetric property of the circle:

$$(3.13)\qquad 4\pi A \leqq \left(\int_{s(x,\lambda)} ds\right)^2 \leqq A_\lambda \int_{s(x,\lambda)} (u_y^2 + u_z^2)^{1/2} ds.$$

Hence

$$(3.14)\qquad A_x^2 + 4\pi A \leqq A_\lambda \int_{s(x,\lambda)} (u_x^2 + u_y^2 + u_z^2)(u_y^2 + u_z^2)^{-1/2} ds,$$

which agrees with the inequality (3.11).

4. It is instructive to compare the proofs given in Section 2 and in this section with the proof suggested by Faber [5] that the capacity is diminished or remains unchanged by the *third* kind of symmetrization defined in the Introduction. This symmetrization replaces the level surface σ_λ ($u = \lambda$) by a sphere σ'_λ of equal volume (with a fixed center) and the "symmetrized" function v is defined as $v = \lambda$ on σ'_λ. Let us denote by $V = V(\lambda)$ and $R = R(\lambda)$ the volume and the radius of this sphere, respectively.

We compare the contribution of the shell between σ_λ and $\sigma_{\lambda+d\lambda}$, $d\lambda > 0$, to the integral $W(u)$ with the corresponding contribution of the spherical shell between σ'_λ and $\sigma'_{\lambda+d\lambda}$ to $W'(v)$. Since the piece of the normal of the surface σ_λ between σ_λ and $\sigma_{\lambda+d\lambda}$ is $|\operatorname{grad} u|^{-1} d\lambda$ we find for the corresponding part of $W(u)$

$$(3.15)\qquad \int\!\!\int |\operatorname{grad} u|^2 \, |\operatorname{grad} u|^{-1} d\lambda \, d\sigma = d\lambda \int_{\sigma_\lambda}\!\!\int |\operatorname{grad} u| \, d\sigma$$

where $d\sigma$ is the area element of σ_λ.

For the symmetrized function v we find

$$(3.16)\qquad |\operatorname{grad} v| = d\lambda/dR = \{R'(\lambda)\}^{-1} = 3\{R(\lambda)\}^{-1} V(\lambda)\{V'(\lambda)\}^{-1}$$

and so the corresponding contribution is

$$(3.17)\qquad d\lambda \cdot 3\{R(\lambda)\}^{-1} V(\lambda)\{V'(\lambda)\}^{-1} \cdot 4\pi\{R(\lambda)\}^2.$$

Thus we have to prove that

$$(3.18)\qquad \int_{\sigma_\lambda}\!\!\int |\operatorname{grad} u|\, d\sigma \geqq (36\pi)^{2/3}\{V(\lambda)\}^{4/3}\{V'(\lambda)\}^{-1}.$$

But by Schwarz's inequality and by the isoperimetric property of the sphere

$$(3.19)\qquad \int_{\sigma_\lambda}\!\!\int |\operatorname{grad} u|\, d\sigma \int_{\sigma_\lambda}\!\!\int |\operatorname{grad} u|^{-1} d\sigma \geqq \Big(\int_{\sigma_\lambda}\!\!\int d\sigma\Big)^2 \geqq (36\pi)^{2/3}\{V(\lambda)\}^{4/3}.$$

The second integral on the left-hand side is $V'(\lambda)$. This proves the assertion.

4. The two-dimensional case. In this section we deal with "two-dimensional condensers" bounded by closed curves s_0 and s_1, s_0 interior to s_1. We shall be concerned in particular with the statements formulated in **1. 4** concerning r_P, $\bar{r}$, C, and Λ.

1. The "two-dimensional capacity" or, shorter, *capacity* of a two-dimensional condenser bounded by s_0 and s_1 can be defined as the minimum of

$$(4.1)\qquad (2\pi)^{-1}\int\!\!\int |\operatorname{grad} \psi|^2 dx\,dy$$

the integration being extended over the set of points between s_0 and s_1; here ψ is an arbitrary function with continuous derivatives defined in the domain of integration and satisfying the conditions: $\psi = 0$ on s_0, $\psi = 1$ on s_1. This capacity differs from the "three-dimensional" capacity of the ring-shaped disk bounded by s_0 and s_1. It is the double of the "capacity per unit height" of a condenser formed by two infinite cylinders; it is related to the concept of the logarithmic potential.

The capacity of a two-dimensional condenser diminishes (or remains unchanged) by symmetrization with respect to a straight line in the plane of the condenser.

The proof follows exactly the same line as in Section **2**.

2. Let s be a closed curve in a plane, P a fixed point in the interior of s. The *inner radius* r_P of s with respect to P can be defined as the radius of the uniquely determined circle around P such that the interior of s can be mapped conformally onto the interior of this circle preserving the point P and the line-element (i. e. the magnitude and direction of the line-element) at P.

If we introduce complex numbers $z = x + iy$ in the plane of s and complex numbers w in the plane of the circle and if P is the point $z = w = 0$ this mapping can be represented in the form

$$w = \phi(z) = z + c_2 z^2 + c_3 z^3 + \cdots, \qquad |w| < r_P, \tag{4.2}$$

where $\phi(z)$ is a power series in z.

Before we discuss the effect of the symmetrization on the inner radius r_P let us consider another formulation of the previous definition. We denote by $G = G(x, y)$ the Green's function of the interior of s with respect to the point $P(0, 0)$ defined by the following properties:

(i) G is harmonic in the interior of s except at P;

(ii) $G = 0$ on s;

(iii) $G = \log(1/r) + g$

where r is the distance from P and $g = g(x, y)$ is regular-harmonic in the interior of s. Then for the inner radius r_P of s with respect to P we obtain

$$\log r_P = g(0, 0). \tag{4.3}$$

Indeed, using the notation (4.2), we have

$$G(x, y) = \log(r_P / |\phi(z)|). \tag{4.4}$$

The definition (4.3) has the advantage that it can be extended without difficulty to the three-dimensional case (in which conformal mapping is not possible in general). See Szegö **20**, p. 587. We now prove the following theorem:

The inner radius r_P of the curve s with respect to an interior point P is not decreased by symmetrizing s with respect to a straight line in its plane.

A similar theorem holds in space (cf. Szegö **20**).

For the proof let us consider the level curve $G = \log(r_P/\eta)$ or $|\phi(z)| = \eta$ where $\eta > 0$, $\eta \to 0$. This level curve obviously lies in the circular ring

$$\eta - a\eta^2 < |z| < \eta + a\eta^2 \tag{4.5}$$

where a is a certain positive constant independent of η. Now let the constant ω be fixed, $\omega > a$. If δ is a sufficiently small positive number and

$$\eta' = \delta - \omega\delta^2, \qquad \eta'' = \delta + \omega\delta^2 \tag{4.6}$$

we have

$$\eta' + a\eta'^2 < \delta < \eta'' - a\eta''^2. \tag{4.7}$$

This means that the circle $|z| = \delta$ is between the two level curves $G = \log(r_P/\eta')$ and $G = \log(r_P/\eta'')$.

But the two-dimensional condenser bounded by the curve s and the level curve $G = \log(r_P/\eta)$ has the capacity[2]

$$\{\log(r_P/\eta)\}^{-1}. \tag{4.8}$$

Hence the condenser bounded by s and the circle $|z| = \delta$ has a capacity which is between the quantities $\{\log(r_P/\eta')\}^{-1}$ and $\{\log(r_P/\eta'')\}^{-1}$ and so has the form

$$\{\log(r_P/\delta) + O(\delta)\}^{-1} \tag{4.9}$$

where $\delta^{-1}O(\delta)$ is bounded as $\delta \to 0$.

After these preliminaries we symmetrize the curve s with respect to a straight line through P. If the resulting curve s' has the inner radius r'_P with respect to the same point P we find, since the circle $|z| = \delta$ is not changed by the symmetrization in question, that

$$\{\log(r_P/\delta) + O(\delta)\}^{-1} \geqq \{\log(r'_P/\delta) + O(\delta)\}^{-1} \tag{4.10}$$

and so $r_P \leqq r'_P$ which is the assertion.

3. The *outer radius* $\bar{r}$ of the curve s can be defined as the radius of the uniquely determined circle onto the exterior of which the exterior of s can be mapped conformally so that the point at infinity and the line element in it are preserved.

If we introduce complex numbers $z = x + iy$ in the plane of s and complex numbers w in the plane of the circle this mapping can be represented in the form

$$w = \chi(z) = z + c_0 + c_1 z^{-1} + c_2 z^{-2} + \cdots; \qquad |w| > \bar{r}. \tag{4.11}$$

A definition of $\bar{r}$ in terms of Green's function is again possible. Moreover the outer radius is identical with the *transfinite diameter* of s [**15**].

Let us consider a condenser with s as interior curve and with a large circle of radius R as exterior curve. Then the capacity will be approximately [cf. **15**, p. 8]

---

[2] This is $(2\pi)^{-1}$ times the flux integral of the harmonic function $\{\log(r_P/\eta)\}^{-1}G$ extended over an arbitrary closed curve inside s and containing the level curve $G = \log(r_P/\eta)$.

$$\{\log(R/\bar{r})\}^{-1}. \tag{4.12}$$

From this it follows as above that *the outer radius of the curve s decreases (or remains unchanged) by symmetrizing s with respect to a straight line in the plane of s.*

4. We consider a cylindrical solid the base of which is bounded by a curve s; the generators may or may not be perpendicular to the bases. Applying the Schwarz symmetrization with an axis perpendicular to the bases of the solid we obtain the following remarkable theorem:

The capacity of a cylindrical solid is never less than the capacity of a right circular cylinder whose base has the same area as that of the given solid.

(The right circular cone has an analogous extremal property.)

In the limiting case in which the height of the cylinder becomes 0 we obtain the statement of 1. 4 concerning the capacity of a disk. Since the capacity of a circular disk with radius 1 is $2/\pi$ [see (8. 10)] we obtain the second inequality in (1. 2).

5. Finally let us consider again the integral (4. 1) extended over the interior of a given curve s. The minimum of this integral under the boundary condition $\psi = 0$ on s and the additional condition

$$\iint \psi^2 dxdy = 1 \tag{4.13}$$

furnishes a constant which can be called the *fundamental frequency* Λ of the disk bounded by s. In discussing the effect of symmetrization on the fundamental frequency of a disk the same argument can be followed as in the case of the capacity. If the original function u satisfies the condition $u = 0$ on s the same will hold for the "symmetrized function" v on the symmetrized curve s'. The normalizing condition (4. 13) will also be preserved. This yields the assertion announced in **1. 4** concerning Λ.

The analogous theorem in space is also true.

5. Proof of the inequalities (1. 4). 1. Let σ_0 be a closed convex surface, σ_1 the outer parallel surface at the distance h, $h > 0$. We denote by $C_h = C$ the capacity of the condenser determined by σ_0 and σ_1. First we establish the lower bound for C given in (1. 4).

This can be easily done by means of the minimum property of the integral (1. 1). In order to perform the integration (1. 1) over the field in the present case we first integrate along the normals of σ_0 and then over σ_0. Let us denote by $\partial\psi/\partial k$ the derivative of ψ in the direction of a fixed normal of σ_0, $0 \leqq k \leqq h$. Then

$$\int_0^h |\operatorname{grad}\psi|^2\,dk \geqq \int_0^h (\partial\psi/\partial k)^2\,dk \tag{5.1}$$

and so, by Schwarz's inequality,

$$1 = \{\int_0^h (\partial\psi/\partial k)\,dk\}^2 \leqq h \int_0^h (\partial\psi/\partial k)^2\,dk; \tag{5.2}$$

hence

$$\int_0^h |\operatorname{grad}\psi|^2\,dk \geqq h^{-1}. \tag{5.3}$$

This yields the lower bound in (1.4).

2. The proof of the upper bound for C given in (1.4) requires a deeper analysis. If $u(h; x, y, z) = u(h; P)$ is the function harmonic in the field between σ_0 and σ_1 for which $u = 0$ on σ_0 and $u = 1$ on σ_1, we have

$$4\pi C_h = \iint (\partial u/\partial n)\,d\sigma \tag{5.4}$$

the integration being extended over an arbitrary closed surface in the field containing σ_0 (n the exterior normal). We show first that C_h, as a function of h, possesses a left-hand derivative $(d/dh)C_h$ and that

$$(d/dh)C_h = -(4\pi)^{-1} \int_{\sigma_1}\!\!\int (\partial u/\partial n)^2 d\sigma. \tag{5.5}$$

3. Let $g = g(x, y, z) = g(P)$ be the function harmonic in the field satisfying the conditions

$$\begin{cases} g = 0 & \text{on } \sigma_0, \\ g = \partial u/\partial n & (n \text{ the interior normal}) \text{ on } \sigma_1. \end{cases} \tag{5.6}$$

The flux-integral

$$4\pi\gamma = \iint (\partial g/\partial n)\,d\sigma \tag{5.7}$$

extended over a closed surface in the field surrounding σ_0 (n the exterior normal) is then independent of the choice of this surface.

We introduce the function

$$\begin{aligned} \bar{u}(x, y, z) &= \bar{u}(P) \\ &= u(h; P) - \epsilon g(P) + C_h^{-1}\gamma\epsilon\{u(h; P) - 1\} - (C_h/C_{h-\epsilon})\,u(h-\epsilon; P) \end{aligned} \tag{5.8}$$

where $\epsilon > 0$. Let $P(x, y, z)$ be an arbitrary point on σ_1 and let PQ be an interior normal-vector of σ_1 of length ϵ so that Q is on the parallel surface σ'_1 at distance $h - \epsilon$ from σ_0. Then at Q we have

$$(5.9)\quad \begin{aligned}\bar{u}(Q) &= u(h;Q) - \epsilon g(Q) + C_h^{-1}\gamma\epsilon\{u(h;Q) - 1\} - C_h/C_{h-\epsilon}\\ &= u(h;P) + \epsilon(\partial u/\partial n)_P - \epsilon g(Q) + C_h^{-1}\gamma\epsilon\{u(h;P) - 1\}\\ &\quad - C_h/C_{h-\epsilon} + O(\epsilon^2) = 1 - C_h/C_{h-\epsilon} + O(\epsilon^2)\end{aligned}$$

(n the interior normal). On the other hand if (x, y, z) is on σ_0 then

$$(5.10)\qquad \bar{u} = -C_h^{-1}\gamma\epsilon.$$

Finally for the flux-integral we have

$$(5.11)\qquad \int\!\!\int (\partial\bar{u}/\partial n)\,d\sigma = 4\pi C_h - 4\pi\gamma\epsilon + C_h^{-1}\gamma\epsilon\cdot 4\pi C_h - 4\pi C_h = 0.$$

But such a function $\bar{u}$ must assume on σ'_1 its maximum and minimum in the ring-shaped domain between σ_0 and σ'_1. Indeed σ_0 is a level surface. Hence if $\bar{u}$ would attain its maximum (minimum) on σ_0 we should have $\partial\bar{u}/\partial n \leqq 0$ ($\geqq 0$) (n the exterior normal). This together with (5.11) would imply that $\partial\bar{u}/\partial n = 0$ which is impossible.

Hence

$$(5.12)\qquad 1 - C_h/C_{h-\epsilon} - a\epsilon^2 < -C_h^{-1}\gamma\epsilon < 1 - C_h/C_{h-\epsilon} + a\epsilon^2$$

where a is a positive constant independent of ϵ. From this we conclude that

$$(5.13)\qquad \lim_{\epsilon\to+0} \frac{C_h - C_{h-\epsilon}}{\epsilon} = \gamma.$$

It is possible to express γ in terms of u; for this purpose we apply Green's formula:

$$(5.14)\qquad \int_{\sigma_0}\!\!\int (g(\partial u/\partial n) - u(\partial g/\partial n))\,d\sigma + \int_{\sigma_1}\!\!\int (g(\partial u/\partial n) - u(\partial g/\partial n))\,d\sigma = 0$$

where the normal is directed into the interior of the field. The first integral vanishes and so

$$(5.15)\qquad \int_{\sigma_1}\!\!\int (\partial u/\partial n)^2 d\sigma = \int_{\sigma_1}\!\!\int (\partial g/\partial n)\,d\sigma = -4\pi\gamma.$$

Combining (5.15) with (5.13) we obtain the fundamental formula (5.5).

4. Let A_h denote the area of the surface σ_1. By using Schwarz's inequality we conclude from (5.4) and (5.5) that

$$(5.16)\qquad (4\pi C_h)^2 \leqq \int_{\sigma_1}\!\!\int (\partial u/\partial n)^2 d\sigma \int_{\sigma_1}\!\!\int d\sigma = -4\pi A_h (d/dh) C_h$$

or

$$(5.17)\qquad (d/dh)(C_h^{-1}) \geqq 4\pi A_h^{-1}.$$

Integrating we obtain, since $\lim C_h^{-1} = 0$ as $h \to +0$,

$$(5.18)\qquad C_h^{-1} \geqq 4\pi \int_0^h A_k^{-1} dk.$$

In the limiting case where σ_1 is a sphere with infinitely large radius this inequality was obtained by Szegö [**21**, p. 423]. The method used in the present proof is very similar to that used in the special case mentioned.

5. Let A and M have the same meanings as in (1.4). It is well known [cf. for instance Szegö **21**, p. 424, (11), (12)] that

$$(5.19)\qquad A_k = A + 2Mk + 4\pi k^2 \leqq A^{-1}(A + Mk)^2;$$

therefore

$$(5.20)\qquad C_h^{-1} \geqq 4\pi A \int_0^h (A + Mk)^{-2} dk = 4\pi h (A + Mh)^{-1}$$

and this establishes the upper estimate of C given in (1.4).

6. Also we have, more precisely,

$$(5.21)\qquad C_h^{-1} \geqq 4\pi \int_0^h (A + 2Mk + 4\pi k^2)^{-1} dk.$$

Performing the integration

$$(5.22)\qquad C_h^{-1} \geqq \frac{2\pi}{\lambda M} \log\left(\frac{(1-\lambda)M + 4\pi h}{(1+\lambda)M + 4\pi h} \cdot \frac{1+\lambda}{1-\lambda}\right); \quad \lambda = (1 - 4\pi A M^{-2})^{1/2}.$$

7. Finally we point out the inequalities corresponding to (1.4) in the two-dimensional case where the capacity is associated with the theory of the logarithmic potential.

Let s_0 be a closed *convex* curve in the plane and s_1 an arbitrary outer curve parallel to s_0. We denote by C the capacity (defined in **4.**1) of the two-dimensional condenser bounded by s_0 and s_1. If L_0 is the length of s_0 and L_1 the length of s_1 we have

$$(5.23) \qquad (L_1/L_0 - 1)^{-1} < C < (\log(L_1/L_0))^{-1}$$

unless s_0 is a circle.

The proof follows exactly the same line of argument as in the case of the Newtonian potential.

6. Capacity of an ellipsoid. 1. Let a, b, c be the semi-axes of an ellipsoid and suppose that

$$(6.1) \qquad a \geqq b \geqq c > 0.$$

Let α, β, γ be the numerical excentricities of the principal sections, each section being perpendicular to the respective axis:

$$(6.2) \qquad \alpha^2 = 1 - c^2/b^2, \quad \beta^2 = 1 - c^2/a^2, \quad \gamma^2 = 1 - b^2/a^2.$$

We denote the capacity of this ellipsoid by $C = C(a, b, c)$.

The following representations of C are well known [cf. for instance E. Heine **6**, p. 156]:

$$(6.3) \qquad C^{-1} = (4\pi)^{-1} \int\int (a^2\xi^2 + b^2\eta^2 + c^2\zeta^2)^{-1/2} d\omega$$

$$(6.4) \qquad = (1/2) \int_0^\infty \{(a^2 + u)(b^2 + u)(c^2 + u)\}^{-1/2} du.$$

The first integral is extended over the unit sphere $\xi^2 + \eta^2 + \zeta^2 = 1$ with the surface element $d\omega$.

2. Two further representations of C^{-1} in the form of infinite series can easily be obtained from (6.4). Straight-forward calculation yields

$$C^{-1} = (2a)^{-1} \int_0^\infty \{(1 + u)(1 - \beta^2 + u)(1 - \gamma^2 + u)\}^{-1/2} du$$

$$(6.5) \qquad = (2a)^{-1} \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} g_\mu g_\nu \beta^{2\mu} \gamma^{2\nu} \int_0^\infty (1 + u)^{-3/2-\mu-\nu} du$$

where we use the abbreviation

$$(6.6) \qquad g_0 = 1; \quad g_m = \frac{1 \cdot 3 \cdots (2m - 1)}{2 \cdot 4 \cdots 2m}, \qquad (m = 1, 2, 3, \cdots).$$

Thus we have the following formula:

$$(6.7) \qquad C^{-1} = a^{-1} \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} \frac{g_\mu g_\nu \beta^{2\mu} \gamma^{2\nu}}{2(\mu + \nu) + 1}$$

$$= a^{-1} \sum_{n=0}^{\infty} \frac{(\beta\gamma)^n}{2n+1} P_n\left(\frac{\beta^2+\gamma^2}{2\beta\gamma}\right) \tag{6.8}$$

where $P_n(t)$ is Legendre's polynomial.

The second formula results from

$$C^{-1} = (2ab)^{-1}c \int_0^\infty \{(1+u)(1+u-\alpha^2 u)(1+u-\beta^2 u)\}^{-\frac{1}{2}} du$$

$$= (2ab)^{-1}c \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} g_\mu g_\nu \alpha^{2\mu}\beta^{2\nu} \int_0^\infty (1+u)^{-3/2-\mu-\nu} u^{\mu+\nu} du$$

$$= (ab)^{-1}c \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} \frac{g_\mu g_\nu}{g_{\mu+\nu}} \frac{\alpha^{2\mu}\beta^{2\nu}}{2(\mu+\nu)+1} \tag{6.9}$$

$$= (ab)^{-1}c \sum_{n=0}^{\infty} \frac{(\alpha\beta)^n}{(2n+1)g_n} P_n\left(\frac{\alpha^2+\beta^2}{2\alpha\beta}\right). \tag{6.10}$$

Here g_n has the same meaning as in (6. 6).

7. Continuation. The following inequalities are known for the capacity of an ellipsoid [Pólya, Szegö **11**; cf. also Szegö **22** and Watson **23**]:

$$\frac{3}{a/(bc)+b/(ca)+c/(ab)} \leqq (abc)^{\frac{1}{3}} \leqq C \leqq \left(\frac{a^2+b^2+c^2}{3}\right)^{\frac{1}{2}} \leqq \frac{bc/a+ca/b+ab/c}{3}. \tag{7.1}$$

We observe that the inequality $(abc)^{\frac{1}{3}} \leqq C$ is equivalent to the special case of the theorem of Poincaré-Faber-Szegö mentioned in **1.** 3 in which σ_0 is an ellipsoid and σ_1 a large sphere.

Here we prove the more precise inequalities (1. 7).

1. We begin with a proof for the upper bound of C. Taking (6. 8) into account we can write the inequality in the following form:

$$\sum_{n=0}^{\infty} \frac{(\beta\gamma)^n}{2n+1} P_n\left(\frac{\beta^2+\gamma^2}{2\beta\gamma}\right) \geqq 3\{1+(1-\beta^2)^{\frac{1}{2}}+(1-\gamma^2)^{\frac{1}{2}}\}^{-1}. \tag{7.2}$$

This is equivalent to the inequality:

$$\sum_{n=0}^{\infty} \frac{h^n}{2n+1} P_n(t) \geqq 3(1+A^{\frac{1}{2}}+B^{\frac{1}{2}})^{-1} \tag{7.3}$$

where

$$\begin{cases} t > 1, & 0 < h < t-(t^2-1)^{\frac{1}{2}} \\ & A = 1-h\{t+(t^2-1)^{\frac{1}{2}}\},\ B = 1-h\{t-(t^2-1)^{\frac{1}{2}}\}. \end{cases} \tag{7.4}$$

Instead of (7. 3) we prove the following stronger result:

$$\frac{\partial}{\partial h}\sum_{n=0}^{\infty}\frac{h^{n+\frac{1}{2}}}{2n+1}P_n(t) \geqq \frac{\partial}{\partial h}\{3h^{\frac{1}{2}}(1+A^{\frac{1}{2}}+B^{\frac{1}{2}})^{-1}\} \tag{7.5}$$

or

$$\sum_{n=0}^{\infty}h^nP_n(t) \geqq -3h(1+A^{\frac{1}{2}}+B^{\frac{1}{2}})^{-2}(A^{-\frac{1}{2}}\partial A/\partial h+B^{-\frac{1}{2}}\partial B/\partial h) + 3(1+A^{\frac{1}{2}}+B^{\frac{1}{2}})^{-1}. \tag{7.6}$$

The left-hand side is

$$(1-2ht+h^2)^{-\frac{1}{2}} = (AB)^{-\frac{1}{2}}. \tag{7.7}$$

Since $h\partial A/\partial h = A-1$, $h\partial B/\partial h = B-1$ the inequality (7. 6) may be written as follows:

$$(AB)^{-\frac{1}{2}} \geqq 3(1+A^{-\frac{1}{2}}+B^{-\frac{1}{2}})(1+A^{\frac{1}{2}}+B^{\frac{1}{2}})^{-2} \tag{7.8}$$

or

$$(1+A^{\frac{1}{2}}+B^{\frac{1}{2}})^2 \geqq 3(A^{\frac{1}{2}}B^{\frac{1}{2}}+A^{\frac{1}{2}}+B^{\frac{1}{2}}). \tag{7.9}$$

This is equivalent to

$$(A^{\frac{1}{2}}-B^{\frac{1}{2}})^2 \geqq -(1-A^{\frac{1}{2}})(1-B^{\frac{1}{2}}). \tag{7.10}$$

2. The proof for the lower bound follows a similar line but the details are more complicated. We prove that

$$\sum_{n=0}^{\infty}\frac{(\beta\gamma)^n}{2n+1}P_n\left(\frac{\beta^2+\gamma^2}{2\beta\gamma}\right) \leqq 3\{(1-\beta^2)^{\frac{1}{4}}(1-\gamma^2)^{\frac{1}{4}}+(1-\beta^2)^{\frac{1}{4}}+(1-\gamma^2)^{\frac{1}{4}}\}^{-1} \tag{7.11}$$

or, in the previous notation,

$$\sum_{n=0}^{\infty}\frac{h^n}{2n+1}P_n(t) \leqq 3(A^{\frac{1}{4}}B^{\frac{1}{4}}+A^{\frac{1}{4}}+B^{\frac{1}{4}})^{-1}. \tag{7.12}$$

Here again we prove more, namely the inequality

$$\frac{\partial}{\partial h}\sum_{n=0}^{\infty}\frac{h^{n+\frac{1}{2}}}{2n+1}P_n(t) \leqq \frac{\partial}{\partial h}\{3h^{\frac{1}{2}}(A^{\frac{1}{4}}B^{\frac{1}{4}}+A^{\frac{1}{4}}+B^{\frac{1}{4}})^{-1}\}. \tag{7.13}$$

Using the previous calculations we can write (7. 13) in the form

$$\begin{aligned}(AB)^{-\frac{1}{2}} \leqq\ & 3(A^{\frac{1}{4}}B^{\frac{1}{4}}+A^{\frac{1}{4}}+B^{\frac{1}{4}})^{-1}\\ & -\tfrac{3}{2}(A^{\frac{1}{4}}B^{\frac{1}{4}}+A^{\frac{1}{4}}+B^{\frac{1}{4}})^{-2}\\ & \times\{A^{-\frac{3}{4}}B^{\frac{1}{4}}(A-1)+A^{\frac{1}{4}}B^{-\frac{3}{4}}(B-1)\\ & \qquad +A^{-\frac{3}{4}}(A-1)+B^{-\frac{3}{4}}(B-1)\},\end{aligned}$$

or

$$\begin{aligned}(7.14)\qquad 2(AB)^{1/4}(A^{1/4}B^{1/4}+A^{1/4}+B^{1/4})^2 &\leqq 6(AB)^{3/4}(A^{1/4}B^{1/4}+A^{1/4}+B^{1/4})\\ &\quad -3(AB)^{3/4}\{A^{-3/4}B^{1/4}(A-1)+A^{1/4}B^{-3/4}(B-1)\\ &\qquad +A^{-3/4}(A-1)+B^{-3/4}(B-1)\}.\end{aligned}$$

Letting $A^{1/4}=r$, $B^{1/4}=s$, this inequality is equivalent to the statement that the polynomial of the seventh degree in r and s

$$(7.15)\qquad f(r,s)=3(r^4s^3+r^3s^4+r^4+s^4+r^3+s^3)-2rs(rs+r+s)^2$$

is *non-negative* for $0\leqq r\leqq 1$, $0\leqq s\leqq 1$. We give two proofs of this assertion.

3. *First proof.* We have

$$\begin{aligned}(7.16)\qquad f(r,s) &= 2(r^2-s^2)(r-s)+3(r^3-s^3)(r-s)\\ &\quad +rs(r-s)^2+2rs(r+s)(1-rs)^2\\ &\quad +2(\tfrac{3}{8}r^3+\tfrac{3}{8}s^3+\tfrac{1}{8}r^4s^3+\tfrac{1}{8}r^3s^4-r^2s^2)\\ &\quad +2(\tfrac{1}{8}r^3+\tfrac{1}{8}s^3+\tfrac{3}{8}r^4s^3+\tfrac{3}{8}r^3s^4-r^3s^3).\end{aligned}$$

The expressions in the last two rows are non-negative since

$$(7.17)\qquad (r^3)^{3/8}(s^3)^{3/8}(r^4s^3)^{1/8}(r^3s^4)^{1/8}=r^2s^2,\quad (r^3)^{1/8}(s^3)^{1/8}(r^4s^3)^{3/8}(r^3s^4)^{3/8}=r^3s^3.$$

Second proof. We show that $f(r,s)$ attains its minimum in the square for $r=s=0$ and for $r=s=1$, this minimum being 0.

Since for $0\leqq r\leqq 1$

$$\begin{aligned}(7.18)\qquad f(r,r) &= 6(r^7+r^4+r^3)-2r^4(r+2)^2\\ &= 2r^3(r-1)^2(3r^2+5r+3)\geqq 0,\\ f(r,0) &= 3(r^4+r^3)\geqq 0\end{aligned}$$

it is sufficient to show that $f(r,s)$ can not attain a relative minimum at an interior point of the square not on the diagonal $r=s$. Now

$$\begin{aligned}(7.19)\qquad f_r(r,s) &= 12r^3s^3+9r^2s^4+12r^3+9r^2\\ &\quad -2s(rs+r+s)^2-4rs(rs+r+s)(1+s),\\ f_s(r,s) &= 12r^3s^3+9s^2r^4+12s^3+9s^2\\ &\quad -2r(rs+r+s)^2-4rs(rs+r+s)(1+r)\end{aligned}$$

and so

$$\begin{aligned}(7.20)\qquad \frac{f_r(r,s)-f_s(r,s)}{r-s} &= -9r^2s^2(r+s)+12(r^2+rs+s^2)+9(r+s)\\ &\quad +2(rs+r+s)^2+4rs(rs+r+s)>0\end{aligned}$$

provided $0 \leqq r < s \leqq 1$ or $0 \leqq s < r \leqq 1$. For such values of r and s the two derivatives $f_r(r, s)$ and $f_s(r, s)$ can not vanish simultaneously.

4. The following parametric representation is due to Watson [**23**]:

$$C = a \operatorname{sn}\omega/\omega, \qquad c = a \operatorname{cn}\omega, \qquad b = a \operatorname{dn}\omega \tag{7.21}$$

where snω, cnω, dnω are the Jacobian elliptic functions, $0 \leqq \omega \leqq K$, and the modulus

$$k = ((a^2 - b^2)/(a^2 - c^2))^{1/2}. \tag{7.22}$$

Thus the main inequalities (1. 7) can be written as follows:

$$\tfrac{1}{3}\{(\operatorname{cn}\omega \operatorname{dn}\omega)^{1/2} + (\operatorname{cn}\omega)^{1/2} + (\operatorname{dn}\omega)^{1/2}\} \leqq \operatorname{sn}\omega/\omega \leqq \tfrac{1}{3}(1 + \operatorname{cn}\omega + \operatorname{dn}\omega), \qquad 0 \leqq \omega \leqq K. \tag{7.23}$$

5. The last inequality in (1. 7) follows easily by use of the representation

$$M = \iint (a^2\xi^2 + b^2\eta^2 + c^2\zeta^2)^{1/2} d\omega \tag{7.24}$$

where the integration is extended over the unit sphere (ξ, η, ζ) with the surface element $d\omega$. Indeed

$$\begin{aligned} a\xi^2 + b\eta^2 + c\zeta^2 &\leqq (a^2\xi^2 + b^2\eta^2 + c^2\zeta^2)^{1/2}(\xi^2 + \eta^2 + \zeta^2)^{1/2} \\ &= (a^2\xi^2 + b^2\eta^2 + c^2\zeta^2)^{1/2}; \end{aligned} \tag{7.25}$$

integration over the unit sphere yields the desired inequality.

8. Ellipsoids of revolution. 1. *Prolate spheroid.* We have $b = c$, $\alpha = 0$, $\beta = \gamma$ so that according to a classical formula

$$A = 2\pi ab((1 - \beta^2)^{1/2} + \beta^{-1} \arcsin \beta); \tag{8.1}$$

on the other hand from (6. 8)

$$C^{-1} = a^{-1}(2\beta)^{-1} \log \frac{1 + \beta}{1 - \beta} \tag{8.2}$$

and $b/a = (1 - \beta^2)^{1/2}$ so that

$$\begin{aligned} (R_A/C)^2 &= A/(4\pi C^2) \\ &= (2\beta)^{-3}(1 - \beta^2)^{1/2}(\beta(1 - \beta^2)^{1/2} + \arcsin \beta)\left(\log \frac{1 + \beta}{1 - \beta}\right)^2 = f(\beta). \end{aligned} \tag{8.3}$$

Now let

$$(8.4) \qquad f(z)=\sum_{n=0}^{\infty} c_{2n}z^{2n}, \qquad c_0=1,\ c_2=c_4=0,\ c_6=-4/945.$$

We prove that *the coefficients* $c_8, c_{10}, c_{12}, \cdots$ *are negative.*

The only singularities of the even function $f(z)$ in the finite part of the complex plane are ± 1. We have

$$(8.5) \qquad c_{2n}=(2\pi i)^{-1}\int f(z)z^{-2n-1}dz$$

where the path of the integration consists of the following parts:

(a) a "small" circle $|z-1|=\epsilon$ enclosing $+1$ in the negative sense;

(b) two straight lines joining $z=1+\epsilon$ and $z=\omega$ along the upper and lower border of the positive real axis;

(c) the "large" semicircle $|z|=\omega$ in the upper half plane;

(d) the curves corresponding to (a), (b), (c) obtained by the transformation $z'=-z$.

The contributions of the circles tend to 0 as $\epsilon \to 0$, $\omega \to \infty$ provided $n \geqq 1$. Now for $1<z<+\infty$

$$(8.6) \quad f(z)=(2z)^{-3}\cdot \mp i(z^2-1)^{1/2}$$
$$\times\left(\mp iz(z^2-1)^{1/2}+\pi/2 \pm i\log(z+(z^2-1)^{1/2})\right)\left(\log\frac{z+1}{z-1}\pm i\pi\right)^2$$

where the upper and lower sign corresponds to the upper and lower border. Here $(z^2-1)^{1/2}$ and the two logarithms are positive.

Hence for $n=1, 2, 3, \cdots$

$$(8.7) \quad c_{2n}=-16^{-1}\int_1^{\infty}(z^2-1)^{1/2}\left\{\left(\log\frac{z+1}{z-1}\right)^2-\pi^2\right.$$
$$\left.+4\log\frac{z+1}{z-1}\left(z(z^2-1)^{1/2}-\log(z+(z^2-1)^{1/2})\right)\right\}z^{-2n-4}dz.$$

Let us write

$$(8.8) \quad \left\{\begin{array}{l} \log\dfrac{z+1}{z-1}=u, \qquad 0<u<+\infty, \\ \log\dfrac{z+1}{z-1}-\pi^2\left(\log\dfrac{z+1}{z-1}\right)^{-1}+4z(z^2-1)^{1/2}-4\log(z+(z^2-1)^{1/2}) \\ =u-\pi^2u^{-1}+4\coth(u/2)(\sinh(u/2))^{-1}-4\log\coth(u/4)=\phi(u). \end{array}\right.$$

Obviously $\lim_{u\to+0}\phi(u) = \lim_{u\to+\infty}\phi(u) = +\infty$. Furthermore

$$(8.9)\qquad (\sinh(u/2))^3\phi'(u) = (\sinh(u/2))^3 + \pi^2 u(u^{-1}\sinh(u/2))^3 - 4$$

is increasing from -4 to $+\infty$. Hence the function $\phi(u)$ is first decreasing and then increasing. Since $c_2 = c_4 = 0$ it can not remain always positive; thus it must change sign exactly twice. But according to a theorem of L. Fejér [cf. Pólya and Szegö **14**, vol. 2, pp. 50-51, problem 81] the sequence $c_2, c_4, c_6, c_8, c_{10}, \cdots$ can not contain more than two vanishing terms, or sign changes. Since $c_2 = c_4 = 0$ and $c_6 < 0$ all following terms must be negative.

At the same time we see that the coefficients of the power series expansion of C/R_A are all *positive.*

2. *Oblate spheroid.* In this case $a = b$, $\alpha = \beta$, $\gamma = 0$ and

$$(8.10)\qquad \begin{cases} A = 2\pi a^2(2\beta)^{-1}\left(2\beta + (1-\beta^2)\log\dfrac{1+\beta}{1-\beta}\right), \\ C^{-1} = a^{-1}\beta^{-1}\operatorname{arc\,sin}\beta \end{cases}$$

so that

$$(8.11)\qquad \begin{aligned}(R_A/C)^2 &= A/(4\pi C^2) \\ &= 2(2\beta)^{-3}(2\beta + (1-\beta^2)\log\frac{1+\beta}{1-\beta})(\operatorname{arc\,sin}\beta)^2 = f_1(\beta).\end{aligned}$$

Let

$$(8.12)\qquad f_1(z) = \sum_{n=0}^{\infty} c'_{2n}z^{2n}, \qquad c'_0 = 1,\ c'_2 = c'_4 = 0,\ c'_6 = 4/945.$$

We prove that *the coefficients* $c'_8, c'_{10}, c'_{12}, \cdots$ *are positive.*

With the same path of integration as in (8. 5) we have

$$(8.13)\qquad c'_{2n} = (2\pi i)^{-1}\int f_1(z)z^{-2n-1}dz.$$

Now for $1 < z < +\infty$

$$(8.14)\qquad \begin{aligned} f_1(z) &= 2(2z)^{-3}(2z + (1-z^2)\log\frac{z+1}{z-1} \pm i\pi(1-z^2)) \\ &\quad\times(\pi/2 \pm i\log(z + (z^2-1)^{1/2}))^2 \end{aligned}$$

on the upper and lower border, respectively. Hence

$$(8.15)\qquad \begin{aligned} c'_{2n} = 4^{-1}\int_1^{\infty}\Big\{&\left(2z + (1-z^2)\log\frac{z+1}{z-1}\right)\log(z + (z^2-1)^{1/2}) \\ &+ (1-z^2)\left(\pi^2/4 - (\log(z + (z^2-1)^{1/2}))^2\right)\Big\}\, z^{-2n-4}dz. \end{aligned}$$

We write

$$(8.16)\quad \begin{cases} z = \cosh u,\ (z^2-1)^{1/2} = \sinh u,\ \log(z+(z^2-1)^{1/2}) = u,\ 0 < u < \\ 2z(z^2-1)^{-1} - \log\dfrac{z+1}{z-1} - (\pi^2/4)\ (\log(z+(z^2-1)^{1/2}))^{-1} \\ \qquad\qquad + \log(z+(z^2-1)^{1/2}) \\ \quad = 2\cosh u(\sinh u)^{-2} - 2\log\coth(u/2) - \pi^2/(4u) + u = \phi_1(u) \end{cases}$$

Again $\lim_{u\to+0}\phi_1(u) = \lim_{u\to+\infty}\phi_1(u) = +\infty$ and

$$(8.17)\qquad (\sinh u)^3\phi'_1(u) = (\sinh u)^3 + (\pi^2/4)u(u^{-1}\sinh u)^3 - 4.$$

Thus we conclude as in the other case that $\phi_1(u)$ changes its sign exactly twice. The rest of the argument is the same as before.

9. Landen's transformation. By means of Landen's transformation we can define certain transformations of a given ellipsoid into another ellipsoid which preserve the capacity C.

1. Let

$$(9.1)\qquad a, b, c;\ \alpha, \beta, \gamma$$

have the same meaning as in Section **6** and let

$$(9.2)\qquad l, m, n;\ \lambda, \mu, \nu$$

be the corresponding quantities for another ellipsoid. Substituting in (6.4)

$$(9.3)\qquad u + a^2 = \beta^2 a^2 (\sin\phi)^{-2}$$

we obtain

$$(9.4)\qquad C^{-1} = (\beta a)^{-1}\int_0^{\Phi}\{1-(\beta^{-1}\gamma)^2\sin^2\phi\}^{-1/2}d\phi$$

where

$$(9.5)\qquad \beta = \sin\Phi,\qquad 0 \leqq \Phi < \pi/2.$$

This is an elliptic integral of the first kind with modulus (7.22). According to Landen's transformation[3] it is equal to

$$(9.6)\quad 2(\beta a)^{-1}(1+\beta^{-1}\gamma)^{-1}\int_0^{\Psi}\{1-4\beta^{-1}\gamma(1+\beta^{-1}\gamma)^{-2}\sin^2\psi\}^{-1/2}d\psi$$

---

[3] See C. Jordan, *Cours d' Analyse,* 3rd Edition, Paris, 1913, vol. 2, pp. 117-118.

where the following relation holds:

(9.7) $$\sin(2\Psi - \Phi) = \beta^{-1}\gamma \sin\Phi = \gamma.$$

We can identify the expression (9.6) with the reciprocal capacity of a second ellipsoid to which the quantities (9.2) correspond provided

(9.8) $$2\mu l = a(\beta + \gamma), \qquad \mu = \sin\Psi, \qquad \mu^{-1}\nu = 2(\beta\gamma)^{1/2}(\beta+\gamma)^{-1}.$$

2. In order to find an explicit representation of l, m, n in terms of a, b, c we set

(9.9) $$\beta = \sin u, \quad \gamma = \sin v, \qquad 0 \leqq v \leqq u < \pi/2,$$

so that from (9.5) and (9.7)

(9.10) $$\Phi = u, \qquad \Psi = (u+v)/2.$$

Consequently in view of (9.8)

(9.11) $$\mu = \sin((u+v)/2), \qquad \nu = (\sin u \sin v)^{1/2}\{\cos((u-v)/2)\}^{-1}.$$

Now from the first equation in (9.8)

(9.12) $$l = a\cos((u-v)/2)$$

and

(9.13) $$\begin{cases} m = l(1-\nu^2)^{1/2} = a\cos((u+v)/[illegible]), \\ n = l(1-\mu^2)^{1/2} = a\cos((u-v)/2)\cos((u+v)/2). \end{cases}$$

To sum up we have found:

(9.14) $$\begin{cases} b = a\cos v, \quad c = a\cos u, \\ l = a\cos((u-v)/2), \quad m = a\cos((u+v)/2), \\ n = a\cos((u-v)/2)\cos((u+v)/2). \end{cases}$$

hus the following representation of our transformation arises:

). 15) $$\begin{cases} 2l = (a+b)^{1/2}(a+c)^{1/2} + (a-b)^{1/2}(a-c)^{1/2}, \\ 2m = (a+b)^{1/2}(a+c)^{1/2} - (a-b)^{1/2}(a-c)^{1/2}, \\ 2n = b + c. \end{cases}$$

3. From (9.14) and from (9.15) we conclude that

. 16) $$l < a, \quad m < b, \quad n > c$$

less $b = c$ in which case $l = a$, $m = b$, $n = c$. Repeated application of th transformation (9.15) leads to a sequence of ellipsoids with the same capacity C to which the quantities

$$a_h, b_h, c_h;\ \alpha_h, \beta_h, \gamma_h \tag{9.17}$$

correspond; (9.17) is identical with (9.1) for $h=0$ and with (9.2) for $h=1$. Since the sequences $\{a_h\}$ and $\{b_h\}$ are decreasing and the sequence $\{c_h\}$ is increasing the limits

$$\lim_{h\to\infty} a_h = a^*, \qquad \lim_{h\to\infty} b_h = b^*, \qquad \lim_{h\to\infty} c_h = c^* \tag{9.18}$$

exist.

Obviously $a \geqq a^* \geqq b^* \geqq c^* \geqq c$. We prove that $b^* = c^*$, that is *the limiting case is always a prolate ellipsoid of revolution.* Indeed let $0 < \gamma < \beta < 1$; from the third equation in (9.8) we have $1 > \mu^{-1}\nu > (\beta^{-1}\gamma)^{1/2}$ so that

$$1 > \beta_h^{-1}\gamma_h > (\beta^{-1}\gamma)^{1/2h}. \tag{9.19}$$

This yields $\lim_{h\to\infty} \beta_h^{-1}\gamma_h = 1$, and shows the rapidity of the convergence.

From (9.15) it is obvious that if the initial ellipsoid is an oblate spheroid the same will be true of all the ellipsoids of the sequence derived from it. In this case the limiting ellipsoid is a sphere and our process reduces to a well known algorithm for the computation of the arc sin function [cf. (8.10)]. Prolate spheroids remain unchanged under the transformation (9.15).

Finally, based on Watson's formulas (7.21) and by the use of the Jacobian elliptic functions, the following representation of the limits (9.18) can be obtained:

$$a^* = \pi C/(2K')\coth(\pi/(2K')),\ b^* = c^* = \pi C/(2K')\{\sinh(\pi/(2K'))\}^{-1}; \tag{9.20}$$

here K' is the complete elliptic integral of the first kind associated with the complementary modulus

$$k' = ((b^2 - c^2)/(a^2 - c^2))^{1/2}. \tag{9.21}$$

4. An interesting property of the transformation (9.15) is the following:

The transformation (9.15) *diminishes the sum and increases the product of the semi-axes, the only exception being the case of a prolate spheroid.*

Using (9.14) we have indeed

$$\left\{\begin{aligned} 1+\cos u+\cos v &\geqq \cos((u+v)/2)+\cos((u-v)/2) \\ &\qquad +\cos((u+v)/2)\cos((u-v)/2), \\ \cos u\cos v &\leqq \{\cos((u+v)/2)\cos((u-v)/2)\}^2 \end{aligned}\right.$$

with equality only for $u = v$.

In view of this remark the proof of the inequalities

$$(abc)^{1/3} \leqq C \leqq \tfrac{1}{3}(a + b + c)$$

could be reduced to the special case of a prolate ellipsoid of revolution. A less simple argument shows that the quantity $(bc)^{1/2} + (ca)^{1/2} + (ab)^{1/2}$ is increased by the transformation (9.15).

10. Further special cases.[4] 1. Let s_0 and s_1 be two closed curves in the x, y-plane, s_1 interior to s_0. We consider an arbitrary continuous function ψ with piece-wise continuous derivatives defined between s_0 and s_1, $\psi = 0$ on s_0 and $\psi = 1$ on s_1. The minimum γ of the integral

$$(4\pi)^{-1} \iint |\operatorname{grad} \psi|^2 dx dy \tag{10.1}$$

the integration being extended over the set of points between s_0 and s_1, is the " capacity per unit length " of a condenser formed by two infinite cylinders S_0 and S_1 whose orthogonal cross-section are s_0 and s_1. (Cf. Section 4.)

We consider a condenser formed by the infinite cylinder S_0 as " exterior surface " and by a *finite* part of S_1, obtained by intersecting the infinite cylinder S_1 orthogonally by two planes at a given distance l from each other. To the finite cylinder we add the two bases so that the " interior surface " is the complete boundary of a finite solid cylinder of height l. The two cylinders have parallel generators. We denote the capacity of this condenser by C_{if}.

A modified form of the previous condenser arises by intersecting also the cylinder S_0 by the planes mentioned, however without considering the bases of the exterior cylinder. In this case the exterior surface is a finite part of a cylinder (without bases) and the interior surface is, as before, the complete boundary of a solid cylinder. We denote the capacity of this condenser by C_{ff}.

Finally we may remove the bases of the interior cylinder and consider the condenser formed by the two finite cylindrical surfaces constructed before, both without the bases. Let C denote the capacity of this condenser.

It is interesting to compare the quantities

$$l\gamma,\ C_{if},\ C_{ff},\ C. \tag{10.2}$$

We prove the following inequalities:

---

[4] Added November 16, 1944.

$$(10.3) \qquad l\gamma < C < C_{ff} < C_{if} < l\gamma + (\gamma A/\pi)^{\frac{1}{2}}$$

where A is the area of the curve s_0.

First we observe that C_{if} is the minimum of the integral

$$(10.4) \qquad (4\pi)^{-1} \int\int\int |\operatorname{grad} \Psi|^2 \, dxdydz$$

where $\Psi(x, y, z)$ is an arbitrary continuous function with piece-wise continuous derivatives in the field between S_0 and the finite cylinder, $\Psi = 0$ on S_0 and $\Psi = 1$ on the finite cylinder. Extending the definition of Ψ to the space exterior to S_0 where we set $\Psi = 0$, we obtain an admissible function for the corresponding minimum problem defining C_{ff}. Finally, extending the definition of any function Ψ^* admissible in this latter problem to the space in the interior of the finite inner cylinder where we set $\Psi^* = 1$, we obtain an admissible function for the minimum problem defining C. This proves the inequalities connecting C_{if}, C_{ff} and C.

As to the lower bound for C we denote by $\Psi_0(x, y, z)$ the function for which the minimum C is attained:

$$(10.5) \qquad C = (4\pi)^{-1} \int\int\int |\operatorname{grad} \Psi_0|^2 \, dxdydz.$$

This integral will be diminished by reducing the infinite domain of integration to the finite domain bounded by the given finite cylinders and the two orthogonal planes mentioned. The integral will be further diminished by omitting from $|\operatorname{grad} \Psi_0|^2$ the term $(\partial\Psi_0/\partial z)^2$ so that

$$(10.6) \qquad C > (1/4\pi) \int_0^l \left\{ \int\int [(\partial\Psi_0/\partial x)^2 + (\partial\Psi_0/\partial y)^2] dxdy \right\} dz.$$

(We assume that the perpendicular cross-sections are $z = 0$ and $z = l$.) For each fixed value of z, $0 < z < l$, the function $\Psi_0(x, y, z)$ is an admissible function for the minimum problem (10.1) so that the interior integral is not less than $4\pi\gamma$. From this the assertion follows.

In order to prove the upper bound for C_{if}, let $\psi_0(x, y)$ denote the function for which the minimum of (10.1) is attained. We define $\psi_0(x, y) = 1$ in the interior of s_1. If $h(z)$ is an arbitrary continuous function of z satisfying the conditions

$$\begin{cases} h(z) = 1 \quad \text{for} \quad 0 \leqq z \leqq l, \\ \lim_{z \to \pm\infty} h(z) = 0, \end{cases}$$

CALCUTTA UNIVERSITY * CENTRAL LIBRARY *

the function $\psi_0(x,y)h(z)$ will be admissible for the minimum problem defining C_{if}. Consequently

$$C_{if} < (1/4\pi)\int_{-\infty}^{\infty}\{\int\int[(\partial\psi_0/\partial x)^2+(\partial\psi_0/\partial y)^2]dxdy\}(h(z))^2dz$$
$$+(1/4\pi)\int_{-\infty}^{\infty}\{\int\int\psi_0{}^2dxdy\}(h'(z))^2dz.$$

In the first integral with respect to x and y we integrate only in the ring-shaped domain between the curves s_0 and s_1; this yields $4\pi\gamma$. The second integral with respect to x and y will be less than A. Hence

$$(10.7)\quad \begin{aligned} C_{if} &< \gamma\int_{-\infty}^{\infty}(h(z))^2dz+(A/4\pi)\int_{-\infty}^{\infty}(h'(z))^2dz \\ &= l\gamma+\gamma\int_{-\infty}^{0}(h(z))^2dz+(A/4\pi)\int_{-\infty}^{0}(h'(z))^2dz \\ &\quad +\gamma\int_{l}^{\infty}(h(z))^2dz+(A/4\pi)\int_{l}^{\infty}(h'(z))^2dz. \end{aligned}$$

Writing, with $\lambda > 0$,

$$(10.8)\qquad \begin{cases} h(z)=e^{\lambda z}, & z\leqq 0,\\ h(z)=e^{-\lambda(z-l)}, & z\geqq l,\end{cases}$$

we obtain

$$(10.9)\qquad C_{if} < l\gamma+\gamma/\lambda+A\lambda/(4\pi).$$

The most advantageous value λ is $(4\pi\gamma/A)^{1/2}$; this furnishes the asserted inequality.

2. Of all closed curves which enclose a given area, the circle has the smallest outer radius (transfinite diameter). This is the area theorem of Bieberbach (see Section **1**).

Of all triangles with given area, the equilateral triangle has the smallest outer radius.

First, the outer radius is a continuous function of the coordinates of the vertices of the triangle. Indeed, let Δ be a given triangle, k a positive constant. We denote by $k\Delta$ the triangle which arises by changing the sides of Δ in the ratio $1:k$. If $\bar{r}$ is the outer radius of Δ, then the outer radius of $k\Delta$ is $k\bar{r}$. Now let ϵ be an arbitrary positive number, $\epsilon < 1$. If the vertices of a triangle Δ_1 are sufficiently near to those of Δ, the triangle Δ_1 can be enclosed between the triangles $(1-\epsilon)\Delta$ and $(1+\epsilon)\Delta$. Hence the assertion follows.

(The same argument shows that the outer radius of an arbitrary closed polygon is a continuous function of the coordinates of the vertices, provided the polygon is convex or, more generally, star-shaped with respect to a certain point.)

Consequently, there must exist a triangle with given area and with a smallest outer radius. If this triangle were not equilateral at least two sides would be different. Symmetrizing with respect to a line perpendicular to the third side, the outer radius would diminish. This is a contradiction so that the triangle with the minimum property must be equilateral.

Of all quadrilaterals with given area, the square has the smallest outer radius.

Let $ABCD$ be a quadrilateral with the smallest outer radius. If two adjacent sides were different, say $AB \neq BC$, we could symmetrize with respect to a line perpendicular to the diagonal AC, and the outer radius would diminish. Hence the minimum can be attained only for quadrilaterals with equal sides. If such a quadrilateral were not a square, symmetrizing with respect to a line perpendicular to a side, we could diminish the outer radius. Consequently, the quadrilateral with the minimum property must be a square.

Of all tetrahedra with given volume, the regular tetrahedron has the smallest capacity.

We symmetrize in this case with respect to a plane perpendicular to one of the edges.

STANFORD UNIVERSITY.

BIBLIOGRAPHY

1. K. Aichi, "Note on the capacity of a nearly spherical conductor and especially of an ellipsoidal conductor," *Proceedings of the Math.-Phys. Society of Tokyo*, series 2, vol. 4 (1908), pp. 243-246.

2. T. Bonnesen and W. Fenchel, "Theorie der konvexen Körper," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 3, Berlin, 1934.

3. R. Clausius, *Die mechanische Behandlung der Electricität*, Braunschweig, 1879.

4. R. Courant, "Beweis des Satzes, dass von allen homogenen Membranen gegebenen Umfanges und gegebener Spannung die kreisförmige den tiefsten Grundton besitzt," *Mathematische Zeitschrift*, vol. 1 (1918), pp. 321-328.

5. G. Faber, "Beweis, dass unter allen homogenen Membranen von gleicher Fläche und gleicher Spannung die kreisförmige den tiefsten Grundton gibt," *Sitzungsberichte der Bayrischen Akademie der Wissenschaften*, 1923, pp. 169-172.

6. E. Heine, *Handbuch der Kugelfunctionen, Theorie und Anwendungen*, 2. edition, Berlin, 1881, vol. 2.

7. E. Krahn, "Über eine von Rayleigh formulierte Minimaleigenschaft des Kreises," *Mathematische Annalen*, vol. 94 (1924), pp. 97-100.

8. E. Krahn, "Über Minimaleigenschaften der Kugel in drei und mehr Dimensionen," *Inaugural-Dissertation*, Göttingen, 1925.

9. J. C. Maxwell, *Electricity and Magnetism*, 3. edition, Oxford, 1892, vol. 1.

10. H. Poincaré, *Figures d'équilibre d'une masse fluide*, Paris, 1903.

11. G. Pólya, Problem, *Archiv der Mathematik un Physik*, series 3, vol. 26 (1917); p. 65; solution by G. Szegö, vol. 28 (1920), pp. 81-82.

12. G. Pólya, "Beitrag zur Verallgemeinerung des Verzerrungssatzes auf mehrfach zusammenhängende Gebiete. Zweite Mitteilung," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1928, pp. 280-282.

13. G. Pólya, "Approximations to the area of the ellipsoid," *Publicaciones del Instituto de Matematica, Rosario*, vol. 5 (1943).

14. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925, vols. 1 and 2.

15. G. Pólya and G. Szegö, "Über den transfiniten Durchmesser (Kapazitätskonstante) von ebenen und räumlichen Punktmengen," *Journal für die reine und angewandte Mathematik*, vol. 165 (1931), pp. 4-49.

16. Lord Rayleigh, "Vibrations of Membranes," *Proceedings of the London Mathematical Society*, vol. 5 (1873), pp. 9-10; *Scientific Papers*, vol. 1, Cambridge, 1899, p. 187.

17. Lord Rayleigh, *The Theory of Sound*, 2. edition, London, vol. 1 (1894).

18. Lord Rayleigh, "On the electrical capacity of approximate spheres and cylinders," *Philosophical Magazine*, vol. 31 (1916), pp. 177-186; *Scientific Papers*, vol. 6, Cambridge, 1920, pp. 383-392.

19. A. Russell, "The eighth Kelvin lecture. Some aspects of Lord Kelvin's life and work," *The Journal of the Institution of Electrical Engineers*, vol. 55 (1917), pp. 1-17.

20. G. Szegö, "Über einige Extremalaufgaben der Potentialtheorie," *Mathematische Zeitschrift*, vol. 31 (1930), pp. 583-593.

21. G. Szegö, "Über einige neue Extremaleigenschaften der Kugel," *Mathematische Zeitschrift*, vol. 33 (1931), pp. 419-425.

22. G. Szegö, "An inequality," *Quarterly Journal of Mathematics*, vol. 6 (1935), pp. 78-79.

23. G. N. Watson, "An inequality," *Quarterly Journal of Mathematics*, vol. 5 (1934), pp. 221-223.

ON FABER POLYNOMIALS.*

By ISSAI SCHUR.[1]

I. Introduction.[2]

Let

$$(1)\quad f(z) = z + a_1 + a_2/z + a_3/z^2 + \cdots = z\sum_{\nu=0}^{\infty} a_\nu z^{-\nu} = zg(1/z),\quad a_0 = 1$$

be a power series concerning the convergence of which no assumption is made.[3]

---

* Received November 8, 1943; Revised February 4, 1944.

[1] Died January 10, 1941, in Tel Aviv, Palestine. The Einstein Institute of Mathematics of the Hebrew University, Jerusalem, has undertaken the complete edition of the posthumous papers of the deceased, its honorary member since 1940. As the realization of this project under present conditions requires considerable time, some of the main results of this scientific legacy will be published in preliminary notes. The present note has been elaborated by Dr. M. Schiffer of the Hebrew University who worked over the notes left on the subject in cooperation with Professor M. Fekete, the general editor of the scientific legacy of the great scholar. The manuscript has been revised in this country.

[2] Grunsky gave necessary and sufficient conditions for the coefficients of a function in order that it be meromorphic and univalent in a given domain D. ("Koeffizientenabschätzungen für schlicht abbildende meromorphe Funktionen," *Mathematische Zeitschrift*, vol. 45 (1939), pp. 29-61). If, in particular, D is the exterior of the unit circle, these conditions take the form

$$\text{(i)}\qquad \Big|\sum_{\mu,\nu=1}^{m} \nu c_{\mu\nu} x_\mu x_\nu\Big| \leqq \sum_{\nu=1}^{m} \nu\, |x_\nu|^2, \qquad (m = 1, 2, \cdots),$$

where the $c_{\mu\nu}$ are defined by the formula (2) of this paper, if the function considered has the form (1). The identity $\nu c_{\mu\nu} = \mu c_{\nu\mu}$ is proved by Grunsky with the aid of Cauchy's residue theorem. The late Professor Schur wanted to bring the conditions (i) into a more easily evaluable form and investigated, therefore, the relations between the coefficients a_ν and the $c_{\mu\nu}$. This paper gives the results he obtained. Another paper, caused by the same problem, dealing with the transformation of quadratic forms to principal axes will appear elsewhere.

[3] In the formal algebra of power series, two series are called equal if corresponding coefficients are identical. We define the sum of $P(x) = \sum_{\nu=a}^{\infty} k_\nu x^\nu$ $(a > -\infty)$ and $P^*(x) = \sum_{\nu=a}^{\infty} k^*_\nu x^\nu$ to be the series $P(x) + P^*(x) = \sum_{\nu=a}^{\infty} (k_\nu + k^*_\nu) x^\nu$ and the product $P(x)P^*(x)$ to be $\sum_{\nu=2a}^{\infty} l_\nu x^\nu$ with $l_\nu = \sum_{\rho=a}^{\nu-a} k_\rho k^*_{\nu-\rho}$. Finally $P(x)^{-1}$ is the power series which satisfies $P(x)P(x)^{-1} = 1$, and the derivative $P'(x)$ of $P(x)$ is $\sum_{\nu=a}^{\infty} \nu k_\nu x^{\nu-1}$.

We define a polynomial $P_m(f)$ in $f(z)$ of degree m $(m=1,2,\cdots)$ such that

$$(2)\quad P_m(f)=z^m+c_{m1}/z+c_{m2}/z^2+\cdots+c_{m\mu}/z^\mu+\cdots=z^m+G_m(1/z),$$

$$G_m(x)=\sum_{\mu=1}^{\infty}c_{m\mu}x^\mu.$$

$P_m(f)$ is called the m-th Faber polynomial of $f(z)$. The existence and uniqueness of $P_m(f)$ for $m\geqq 1$ is easily shown by recursion.

Let

$$(3)\qquad Q(f)=q_0z^m+q_1z^{m-1}+\cdots+q_m+q'/z+\cdots$$

be any polynomial in $f(z)$ of degree m. Then, writing $P_0(f)=1$,

$$D(f)=Q(f)-q_0P_m(f)-q_1P_{m-1}(f)-\cdots-q_mP_0(f)=\alpha/z+\cdots$$

is a polynomial in $f(z)$ the development of which with respect to z contains only negative powers. This being evidently impossible unless $D(f)$ is identically zero, we have the development

$$(3')\qquad Q(f)=q_0P_m(f)+q_1P_{m-1}(f)+\cdots+q_mP_0(f).$$

Letting

$$(4)\qquad g(x)^m=\sum_{\mu=0}^{\infty}a_{m\mu}x^\mu \qquad (m=1,2,\cdots),\quad a_{m0}=1$$

and writing $x=1/z$ we have

$$f(z)^m=z^mg(x)^m=z^m+a_{m1}z^{m-1}+a_{m2}z^{m-2}+\cdots+a_{mm}+a_{m,m+1}/z+\cdots$$

whence, according to (3) and (3′),

$$(5)\quad f(z)^m=P_m(f)+a_{m1}P_{m-1}(f)+\cdots+a_{m,m-1}P_1(f)+a_{mm}P_0(f).$$

Let $\phi_m(x)=1+a_{m1}x+\cdots+a_{mm}x^m$ and $\psi_m(x)=a_{m,m+1}x+\cdots+a_{m,m+\nu}x^\nu+\cdots$. Then $f(z)^m=z^m\phi_m(x)+\psi_m(x)$ and therefore, by (2) and (5),

$$(6)\qquad \psi_m(x)=G_m(x)+a_{m1}G_{m-1}(x)+\cdots+a_{m,m-1}G_1(x).$$

This important identity establishes a relation between the coefficients $c_{\mu\nu}$ defined in (2) and the $a_{\mu\nu}$ defined in (4). In fact, comparing coefficients of like powers of x, we have for $\nu\geqq 1$, $m\geqq 1$,

$$(7)\qquad a_{m,m+\nu}=c_{m\nu}+a_{m1}c_{m-1,\nu}+a_{m2}c_{m-2,\nu}+\cdots+a_{m,m-1}c_{1\nu}.$$

In order to combine all these formulas in one, we introduce the infinite matrices

$$(8)\quad \begin{cases} A = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ a_{21} & 1 & 0 & \cdots \\ a_{32} & a_{31} & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} = (a_{\mu,\mu-\nu}), \quad a_{\mu 0} = 1, \quad a_{\mu,-k} = 0 \text{ for } k \geqq 1, \\ B = \begin{pmatrix} a_{12} & a_{13} & \cdots \\ a_{23} & a_{24} & \cdots \\ a_{34} & a_{35} & \cdots \\ \cdot & \cdot & \cdot \end{pmatrix} = (a_{\mu,\mu+\nu}), \qquad C = \begin{pmatrix} c_{11} & c_{12} & \cdots \\ c_{21} & c_{22} & \cdots \\ \cdot & \cdot & \cdot \end{pmatrix} = (c_{\mu\nu}). \end{cases}$$

Then (7) can be expressed in the equivalent forms

$$(7') \qquad B = AC, \qquad C = A^{-1}B.$$

With the aid of (7′) we shall give an *explicit formula for the* $c_{\mu\nu}$ *in terms of the coefficients* a_ν *of* $f(z)$. We shall see that each $c_{\mu\nu}$ is a polynomial in the a_ν with non-negative integer coefficients, and that $\nu c_{\mu\nu} = \mu c_{\nu\mu}$ (Grunsky's identity). This can also be shown by other arguments [4] but we shall calculate the coefficients of these polynomials explicitly, and shall see in particular that Grunsky's formula is an expression of a corresponding symmetry property of the polynomial coefficients.

II. Computation of the elements of the matrix *C*. We define, in conformity with (4),

$$(4') \qquad g(x)^{-m} = \sum_{\mu=0}^{\infty} a_{-m,\mu} x^{\mu}, \qquad (m = 1, 2, \cdots), \; a_{-m,0} = 1.$$

In particular, we have in $a_{-1,\mu} = p_\mu$ the well-known Aleph-functions of Wronski. In order to establish relations between the $a_{-m,\mu}$ and the $a_{n\mu}$, we make use of the following simple lemma:

LEMMA. *Let* $g(x) = \sum_{\nu=0}^{\infty} a_\nu x^\nu$ *be an arbitrary power series. Then*

$$[g(x)^k - xg'(x)g(x)^{k-1}]_k = 0$$

where $[u(x)]_k$ *denotes the coefficient of* x^k *in the development of* $u(x)$ *in powers of* x.

---

[4] The integral character of the coefficients follows immediately by induction from (7) since i) $a_{m,\mu}$ (by (1) and (4)) is a polynomial in a_ν with integral coefficients for $m \geqq 1$, $\mu \geqq 0$; ii) $c_{1\nu} = a_{\nu+1}$ for $\nu \geqq 1$ by (1) and (2).

The truth of the lemma is evident since

$$g(x)^k = \sum_{\rho=0}^{\infty} a_{k\rho} x^\rho \quad \text{and} \quad xg'(x)\,g(x)^{k-1} = \sum_{\rho=0}^{\infty} (\rho/k)\,a_{k\rho} x^\rho.$$

We apply the lemma with $k = \mu - \nu$, μ and ν $(\nu < \mu)$ being arbitrary positive integers, and obtain

$$0 = [g(x)^{\mu-\nu} - xg'(x)\,g(x)^{\mu-\nu-1}]_{\mu-\nu} = \left[g(x)^\mu\left(\frac{1}{g(x)^\nu} - \frac{xg'(x)}{g(x)^{\nu+1}}\right)\right]_{\mu-\nu}$$
$$= \left[\frac{g(x)^\mu}{\nu x^{\nu-1}}\left(\frac{x^\nu}{g(x)^\nu}\right)'\right]_{\mu-\nu} = \left[\frac{g(x)^\mu}{\nu}\sum_{\lambda=0}^{\infty}(\lambda+\nu)\,a_{-\nu,\lambda}x^\lambda\right]_{\mu-\nu}.$$

Hence

$$(9) \qquad a_{\mu,\mu-\nu} + \frac{\nu+1}{\nu}\,a_{\mu,\mu-\nu-1}a_{-\nu,1} + \frac{\nu+2}{\nu}\,a_{\mu,\mu-\nu-2}a_{-\nu,2} + \cdots + \frac{\mu}{\nu}\,a_{-\nu,\mu-\nu} = 0,$$

which, by (8) and (4′), yields

$$(10) \qquad A^{-1} = \left(\frac{\mu}{\nu}\,a_{-\nu,\mu-\nu}\right), \qquad a_{-\nu,-k} = 0 \text{ for } k \geqq 1.$$

From (7′) and (10) we obtain the formula

$$(11) \qquad c_{\mu\nu} = \sum_{\lambda=1}^{\mu} \frac{\mu}{\lambda}\,a_{-\lambda,\mu-\lambda}a_{\lambda,\lambda+\nu}$$

as a starting point for further calculations.

We begin by computing $a_{-1,\mu} = p_\mu$, for which we obtain the well-known formula

$$(12) \qquad p_\mu = \sum(-1)^{a_1+a_2+\cdots+a_\mu}\frac{(\alpha_1+\alpha_2+\cdots+\alpha_\mu)!}{\alpha_1!\,\alpha_2!\cdots\alpha_\mu!}\,a_1^{a_1}\cdots a_\mu^{a_\mu}$$
$$(\alpha_1 + 2\alpha_2 + \cdots + \mu\alpha_\mu = \mu).$$

Differentiating the identity $g(x)^{-1} = \sum_{\mu=0}^{\infty} p_\mu x^\mu$ $\lambda - 1$ times with respect to a_1 we have

$$(13) \qquad (-1)^{\lambda-1}\frac{(\lambda-1)!\,x^{\lambda-1}}{g(x)^\lambda} = \sum_{\mu=0}^{\infty}\frac{\partial^{\lambda-1}p_\mu}{\partial a_1{}^{\lambda-1}}\,x^\mu.$$

Hence by (4′)

$$(14) \qquad a_{-\lambda,\mu-\lambda} = (-1)^{\lambda-1}\frac{1}{(\lambda-1)!}\,\frac{\partial^{\lambda-1}p_{\mu-1}}{\partial a_1{}^{\lambda-1}}$$

and so by (12)

$$(15) \qquad a_{-\lambda,\mu-\lambda} = \sum_{a_1+2a_2+\cdots+(\mu-1)a_{\mu-1}=\mu-1}(-1)^{\lambda-1+a_1+\cdots+a_{\mu-1}}$$
$$\times\frac{(\alpha_1+\alpha_2+\cdots\alpha_{\mu-1})!}{\alpha_1!\,\alpha_2!\cdots\alpha_{\mu-1}!}\binom{\alpha_1}{\lambda-1}a_1{}^{a_1-(\lambda-1)}a_2{}^{a_2}\cdots a_{\mu-1}{}^{a_{\mu-1}}.$$

The m-th Faber polynomial $P^*_m(f^*)$ of $f^*(z) = f(z) + c$ is evidently connected with the m-th Faber polynomial $P_m(f)$ of $f(z)$ by the relation $P^*_m(f^*) = P_m(f^* - c)$ which, since $P_m(f^* - c) = P_m(f)$, shows that the matrices C associated according to (2) and (8) with $f(z)$ and $f^*(z)$ are the same and thus do not depend on a_1. For our final aim, to compute the elements $c_{\mu\nu}$ of C, we may, therefore, assume henceforth that $a_1 = 0$. The coefficients which correspond to this assumption will be denoted $a_{ik}{}^{(0)}$.

From (15) (with $a_1 = 0$) we have

$$(16) \quad a^{(0)}_{-\lambda,\mu-\lambda} = \sum (-1)^{\alpha_2+\alpha_3+\cdots+\alpha_{\mu-\lambda}} \frac{(\lambda - 1 + \alpha_2 + \cdots \alpha_{\mu-\lambda})!}{(\lambda-1)!\,\alpha_2!\cdots\alpha_{\mu-\lambda}!} a_2{}^{\alpha_2}\cdots a_{\mu-\lambda}^{\alpha_{\mu-\lambda}}$$

$$(2\alpha_2 + \cdots + (\mu-\lambda)\alpha_{\mu-\lambda} = \mu - \lambda).$$

Also

$$(17) \quad a^{(0)}_{\lambda,\mu+\lambda} = \sum \frac{\lambda!}{(\lambda - \beta_2 - \beta_3 - \cdots)!\,\beta_2!\beta_3!\cdots\beta_{\mu+\lambda}!} a_2{}^{\beta_2} a_3{}^{\beta_3}\cdots a_{\mu+\lambda}^{\beta_{\mu+\lambda}}$$

$$(2\beta_2 + \cdots + (\mu+\lambda)\beta_{\mu+\lambda} = \mu + \lambda).$$

Introducing (16) and (17) into (11) we get

$$(18) \quad c_{\mu\nu} = \sum_{\lambda=1}^{\mu} (\mu/\lambda) \sum_{A=\mu-\lambda} (-1)^{\alpha} \frac{(\lambda - 1 + \alpha)!\,a_2{}^{\alpha_2}a_3{}^{\alpha_3}\cdots a_{\mu-\lambda}^{\alpha_{\mu-\lambda}}}{(\lambda-1)!\,\alpha_2!\,\alpha_3!\cdots\alpha_{\mu-\lambda}!}$$

$$\times \sum_{B=\lambda+\nu} \frac{\lambda!\,a_2{}^{\beta_2}a_3{}^{\beta_3}\cdots a_{\lambda+\nu}^{\beta_{\lambda+\nu}}}{(\lambda-\beta)!\,\beta_2!\,\beta_3!\cdots\beta_{\lambda+\nu}!}$$

where the abbreviations $\alpha = \alpha_2 + \alpha_3 + \cdots + \alpha_{\mu-\lambda}$, $\beta = \beta_2 + \beta_3 + \cdots + \beta_{\lambda+\nu}$, $A = 2\alpha_2 + 3\alpha_3 + \cdots + (\mu-\lambda)\alpha_{\mu-\lambda}$, $B = 2\beta_2 + 3\beta_3 + \cdots + (\lambda+\nu)\beta_{\lambda+\nu}$ have been introduced. From (18) we see that $c_{\mu\nu}$ has degree $[\frac{1}{2}(\mu+\nu)]$ (at most) and weight $\mu + \nu$.

Let

$$(19) \quad c_{\mu\nu} = \sum_{\Gamma=\mu+\nu} C^{(\mu,\nu)}_{\gamma_2\gamma_3\cdots\gamma_{\mu+\nu}} a_2{}^{\gamma_2}a_3{}^{\gamma_3}\cdots a_{\mu+\nu}^{\gamma_{\mu+\nu}},$$

$$\gamma = \gamma_2 + \cdots + \gamma_{\mu+\nu}, \quad \Gamma = 2\gamma_2 + \cdots + (\mu+\nu)\gamma_{\mu+\nu}.$$

We have now to compute the integers $C^{(\mu,\nu)}_{\gamma_2\gamma_3\cdots\gamma_{\mu+\nu}}$. From (18) and (19) we obtain

$$(20) \quad C^{(\mu,\nu)}_{\gamma_2\gamma_3\cdots\gamma_{\mu+\nu}} = \sum_{\lambda=1}^{\mu} \mu \sum_{A=\mu-\lambda} (-1)^{\alpha} \frac{(\lambda-1+\alpha)!}{\gamma_2!\cdots\gamma_{\mu+\nu}!\,(\lambda-\gamma+\alpha)!}$$

$$\times \binom{\gamma_2}{\alpha_2}\cdots\binom{\gamma_{\mu-\lambda}}{\alpha_{\mu-\lambda}} = \frac{\mu(\gamma-1)!}{\gamma_2!\cdots\gamma_{\mu+\nu}!} \sum_{\lambda=1}^{\mu} \sum_{A=\mu-\lambda} (-1)^{\alpha} \binom{\lambda-1+\alpha}{\gamma-1}\binom{\gamma_2}{\alpha_2}\cdots ($$

Taking into consideration that in (20) the summation indices $\lambda, \alpha_2, \cdots, \alpha_{\mu+\nu}$ are always connected by the equation $\lambda = \mu - A$, we may transform it into the form

$$(21)\quad C^{(\mu,\nu)}_{\gamma_2\gamma_3\cdots\gamma_{\mu+\nu}} = \frac{\mu(\gamma-1)!}{\gamma_2!\,\gamma_3!\cdots\gamma_{\mu+\nu}!} \Sigma\,(-1)^{\alpha}\binom{\gamma_2}{\alpha_2}\binom{\gamma_3}{\alpha_3}\cdots\binom{\gamma_{\mu-1}}{\alpha_{\mu-1}}\binom{\mu-1-A+\alpha}{\gamma-1}$$

where the summation is to be extended over all non-negative integer values of α_i, the symbol $\binom{u}{v}$ being defined in the usual way for $u \geqq v$ and as 0 for $u < v$ even if u is negative. Thus we have to calculate only the expressions

$$(22)\quad D^{(\mu)}_{\gamma_2\gamma_3\cdots\gamma_{\mu+\nu}} = \Sigma\,(-1)^{\alpha}\binom{\gamma_2}{\alpha_2}\binom{\gamma_3}{\alpha_3}\cdots\binom{\gamma_{\mu-1}}{\alpha_{\mu-1}}\binom{\mu-1-A+\alpha}{\gamma-1}.$$

Since (with our convention concerning $\binom{u}{v}$) the expression $\binom{\mu-1-A+\alpha}{\gamma-1}$ vanishes unless $\mu - A + \alpha \geqq \gamma$, that is unless $\mu - \gamma \geqq \alpha_2 + 2\alpha_3 + \cdots + (\mu+\nu-1)\alpha_{\mu+\nu}$, we see that $\alpha_\mu = \alpha_{\mu+\nu} = \cdots = \alpha_{\mu+\nu} = 0$, and so we have

$$(23)\quad D^{(\mu)}_{\gamma_2\cdots\gamma_{\mu+\nu}} = \Sigma\,(-1)^{\alpha}\binom{\gamma_2}{\alpha_2}\cdots\binom{\gamma_{\mu+\nu}}{\alpha_{\mu+\nu}}\binom{\mu-1-A+\alpha}{\gamma-1}$$

where α_i again takes only non-negative integer values and

$$\alpha = \alpha_2 + \alpha_3 + \cdots + \alpha_{\mu+\nu}, \qquad A = 2\alpha_2 + 3\alpha_3 + \cdots + (\mu+\nu)\alpha_{\mu+\nu}.$$

III. The explicit formula for $c_{\mu\nu}$. The expression (23) can be summed successively with the aid of the following lemma.

LEMMA. *Let m and n be integers, $m \geqq 1$, $n \geqq 0$. Let $b^{(m)}_{n,k} = 0$ for $k > n(m-1)$ or $k < 0$, and let $b^{(m)}_{n,k}$ be defined for $0 \leqq k \leqq n(m-1)$ by*

$$(24)\qquad \left(\frac{1-x^m}{1-x}\right)^n \equiv \sum_{k=0}^{n(m-1)} b^{(m)}_{n,k}\, x^k.$$

(Thus $b^{(m)}_{n,k}$ is a non-negative integer for $m \geqq 1$, $n \geqq 0$, and k arbitrary). *Then* (assuming the above convention concerning $\binom{u}{v}$) *we have for arbitrary positive integers h and r the identities*

$$(25)\quad \sum_{\nu=0}^{n}(-1)^{\nu}\binom{n}{\nu}\binom{h+n-1+r-m\nu}{h+n-1}=\sum_{\rho=0}^{r}b_{n,\rho}^{(m)}\binom{h-1+r-\rho}{h-1}$$
$$=\sum_{\rho=-\infty}^{+\infty}b_{n,\rho}^{(m)}\binom{h-1+r-\rho}{h-1}.$$

((25) is trivially true for $h>0$, $r\leqq 0$).

We have (by the binomial theorem)

$$(26)\qquad (1-x)^{-(h+n)}=\sum_{\mu=0}^{\infty}\binom{h+n-1+\mu}{h+n-1}x^{\mu}$$

whence

$$(27)\qquad \left[\frac{(1-x^{m})^{n}}{(1-x)^{h+n}}\right]_{r}=\sum_{\mu+m\nu=r}^{\mu,\nu\geqq 0}(-1)^{\nu}\binom{n}{\nu}\binom{h+n-1+\mu}{h+n-1}$$
$$=\sum_{\nu=0}^{n}(-1)^{\nu}\binom{n}{\nu}\binom{h+n-1+r-m\nu}{h+n-1}.$$

On the other hand, by (24),

$$(28)\qquad \left[\frac{(1-x^{m})^{n}}{(1-x)^{h+n}}\right]_{r}=\left[\sum_{\rho=0}^{n(m-1)}b_{n,\rho}^{(m)}x^{\rho}\sum_{\sigma=0}^{\infty}\binom{h-1+\sigma}{h-1}x^{\sigma}\right]_{r}$$
$$=\sum_{\rho+\sigma=r}^{\rho,\sigma\geqq 0}b_{n,\rho}^{(m)}\binom{h-1+\sigma}{h-1}=\sum_{\rho=0}^{r}b_{n,\rho}^{(m)}\binom{h-1+r-\rho}{h-1}.$$

Comparing (27) and (28) we obtain (25).

In the case $h=0$, $n>0$, combination of (24) with (27) yields the additional equality

$$(29)\qquad \sum_{\nu=0}^{n}(-1)^{\nu}\binom{n}{\nu}\binom{n-1+r-m\nu}{n-1}=b_{n,r}^{(m)}.$$

To carry out the summation in (23) we apply (25) and (29). Let γ_j $(2\leqq j\leqq \mu+\nu)$ be the last non-vanishing term in $\gamma_2,\cdots,\gamma_{\mu+\nu}$. If $j=2$, then by (23) and (29) (and the convention about $\binom{u}{v}$)

$$(30)\quad D_{\gamma_2\cdots\gamma_{\mu+\nu}}^{(\mu)}=\sum_{\alpha_2}(-1)^{\alpha_2}\binom{\gamma_2}{\alpha_2}\binom{\mu-1-\alpha_2}{\gamma_2-1}=b_{\gamma_2,\mu-\gamma_2}^{(1)}.$$

If $j\geqq 3$ we set $\alpha=\alpha_3+\cdots+\alpha_j$, $\bar{A}=3\alpha_3+\cdots+j\alpha_j$, $\bar{\gamma}=\gamma_3+\cdots+\gamma_j$, and obtain

$$(31)\quad D_{\gamma_2\cdots\gamma_{\mu+\nu}}^{(\mu)}=\sum_{\alpha_3,\ldots,\alpha_j}(-1)^{\bar{\alpha}}\binom{\gamma_3}{\alpha_3}\cdots\binom{\gamma_j}{\alpha_j}$$
$$\sum_{\alpha_2}(-1)^{\alpha_2}\binom{\gamma_2}{\alpha_2}\binom{\mu-1-\bar{A}+\bar{\alpha}-\alpha_2}{\bar{\gamma}+\gamma_2-1}.$$

The inner sum can be evaluated by applying (25) with $\nu = \alpha_2$, $n = \gamma_2$, $h = \bar{\gamma}$, $m = 1$, $r = \mu - \gamma - \bar{A} + \bar{\alpha}$. We obtain

$$(32) \quad \sum_{\alpha_2=0}^{\gamma_2} (-1)^{\alpha_2} \binom{\gamma_2}{\alpha_2} \binom{\mu - 1 - \bar{A} + \bar{\alpha} - \alpha_2}{\bar{\gamma} + \gamma_2 - 1} = \sum_{\rho} b^{(1)}_{\gamma_2,\rho} \binom{\mu - \gamma_2 - \rho - 1 - \bar{A} + \bar{\alpha}}{\bar{\gamma} - 1}$$

and thus from (23) and (31)

$$(33) \qquad D^{(\mu)}_{\gamma_2 \cdots \gamma_{\mu+\nu}} = \sum_{\rho} b^{(1)}_{\gamma_2,\rho} D^{(\mu-\gamma_2-\rho)}_{0\gamma_3 \cdots \gamma_{\mu+\nu}} .$$

Now if $j \geqq 4$ we separate all terms in $D^{(\mu-\gamma_2-\rho)}_{0\gamma_3 \cdots \gamma_{\mu+\nu}}$ which contain α_3 and apply (25) with $\nu = \alpha_3$, $n = \gamma_3$, and $m = 2$, obtaining

$$(34) \qquad D^{(\mu)}_{\gamma_2 \cdots \gamma_{\mu+\nu}} = \sum_{\rho_2,\rho_3} b^{(1)}_{\gamma_2,\rho_2} b^{(2)}_{\gamma_3,\rho_3} D^{(\mu-\gamma_2-\gamma_3-\rho_2-\rho_3)}_{00\gamma_4 \cdots \gamma_{\mu+\nu}} .$$

We continue in this way and at each step the dependence of $D^{(\mu')}_{\gamma_2 \cdots \gamma_{\mu+\nu}}$ on a further $b^{(i-1)}_{\gamma_i,\rho_i}$ is expressed. Finally we consider $D^{(\mu-\gamma_2-\cdots-\gamma_{j-1}-\rho_2-\cdots-\rho_{j-1})}_{00\cdots0\gamma_j\cdots\gamma_{\mu+\nu}}$. Let $\mu' = \mu - \gamma_2 - \cdots - \gamma_{j-1} - \rho_2 - \cdots \rho_{j-1}$. Then by (23) and (29) (with $\alpha_j = \nu$, $n = \gamma_j$, $r = \mu' - \gamma_j$, $m = j - 1$) we have

$$(35) \quad D^{(\mu')}_{00\cdots0\gamma_j\cdots\gamma_{\mu+\nu}} = \sum (-1)^{\alpha_j} \binom{\gamma_j}{\alpha_j} \binom{\mu' - 1 - (j-1)\alpha_j}{\gamma_j - 1} = b^{(j-1)}_{\gamma_j,\mu-\gamma-\rho_2-\cdots-\rho_{j-1}} .$$

Hence if $j \geqq 3$

$$(36) \quad D^{(\mu)}_{\gamma_2 \cdots \gamma_{\mu+\nu}} = \sum_{\rho_2,\cdots,\rho_{j-1}} b^{(1)}_{\gamma_2,\rho_2} b^{(2)}_{\gamma_3,\rho_3} \cdots b^{(j-2)}_{\gamma_{j-1},\rho_{j-1}} b^{(j-1)}_{\gamma_j,\mu-\gamma-\rho_2-\cdots-\rho_{j-1}} .$$

Since $b^{(m)}_{0,k} = 0$ for $k > 0$ and $b^{(m)}_{0,k} = 1$ for $k = 0$, we may write (30) and (36) in the common form (valid for each admissible set of the γ_i's)

$$(37) \quad D^{(\mu)}_{\gamma_2 \cdots \gamma_{\mu+\nu}} = \sum b^{(1)}_{\gamma_2,\rho_2} b^{(2)}_{\gamma_3,\rho_3} \cdots b^{(\mu+\nu-1)}_{\gamma_{\mu+\nu},\rho_{\mu+\nu}} ; \quad \rho_2 + \rho_3 + \cdots + \rho_{\mu+\nu} = \mu - \gamma .$$

In particular we see by (19), (21), (22) and (37) that the $c_{\mu\nu}$ are *polynomials in* $a_2, a_3, \cdots$ *with non-negative integer coefficients.*

An elegant expression can be given to (37) in the following way. By the definition of the $b^{(m)}_{n,k}$ we have

$$\left(\frac{1 - x^{\lambda-1}}{1 - x} \right)^{\gamma_\lambda} = \sum_{\rho_\lambda=0}^{\infty} b^{(\lambda-1)}_{\gamma_\lambda,\rho_\lambda} x^{\rho_\lambda}$$

and so

$$\left[\prod_{\lambda=2}^{\mu+\nu} \left(\frac{1 - x^{\lambda-1}}{1 - x} \right)^{\gamma_\lambda} \right]_{\mu-\gamma} = \sum_{\rho_2,\cdots,\rho_{\mu+\nu}} b^{(1)}_{\gamma_2,\rho_2} b^{(2)}_{\gamma_3,\rho_3} \cdots b^{(\mu+\nu-1)}_{\gamma_{\mu+\nu},\rho_{\mu+\nu}} .$$

Hence we have the following

THEOREM. *Let* $\gamma = \gamma_2 + \gamma_3 + \cdots + \gamma_{\mu+\nu}$, $\Gamma = 2\gamma_2 + 3\gamma_3 + \cdots + (\mu+\nu)\gamma_{\mu+\nu}$. *Then*

$$(38) \quad c_{\mu\nu} = \sum_{\Gamma=\mu+\nu} \frac{\mu(\gamma-1)!}{\gamma_2!\,\gamma_3!\cdots\gamma_{\mu+\nu}!} \left[\prod_{\lambda=2}^{\mu+\nu} \left(\frac{x-x^{\lambda}}{1-x} \right)^{\gamma_\lambda} \right]_{\mu} a_2^{\gamma_2} a_3^{\gamma_3} \cdots a_{\mu+\nu}^{\gamma_{\mu+\nu}},$$

where $[\cdots]_\mu$ *denotes the* μ*-th coefficient of* x *in the expansion of the expression in the bracket.*

Grunsky's law of symmetry, namely that $\nu c_{\mu\nu} = \mu c_{\nu\mu}$, may be derived immediately from (38). For

$$\nu C^{(\mu,\nu)}_{\gamma_2\cdots\gamma_{\mu+\nu}} = \mu\nu \frac{(\gamma-1)!}{\gamma_2!\cdots\gamma_{\mu+\nu}!} D^{(\mu)}_{\gamma_2\cdots\gamma_{\mu+\nu}}$$

and

$$\mu C^{(\nu,\mu)}_{\gamma_2\cdots\gamma_{\mu+\nu}} = \mu\nu \frac{(\gamma-1)!}{\gamma_2!\cdots\gamma_{\mu+\nu}!} D^{(\nu)}_{\gamma_2\cdots\gamma_{\mu+\nu}}.$$

Therefore, we have only to prove that

$$(39) \qquad \left[\prod_{\lambda=2}^{\mu+\nu} \left(\frac{1-x^{\lambda-1}}{1-x} \right)^{\gamma_\lambda} \right]_{\mu-\gamma} = \left[\prod_{\lambda=2}^{\mu+\nu} \left(\frac{1-x^{\lambda-1}}{1-x} \right)^{\gamma_\lambda} \right]_{\nu-\gamma}.$$

Now the expression

$$(40) \quad q(x) = \prod_{\lambda=2}^{\mu+\nu} \left(\frac{1-x^{\lambda-1}}{1-x} \right)^{\gamma_\lambda} = 1 + q_1 x + \cdots + q_n x^n, \quad n = \sum_{\lambda=2}^{\mu+\nu} (\lambda-2)\gamma_\lambda$$

satisfies the equation $x^n q(1/x) = q(x)$ which yields $q_{n-s} = q_s$. Since $n = \Gamma - 2\gamma = (\mu-\gamma) + (\nu-\gamma)$ we thus have $q_{\mu-\gamma} = q_{\nu-\gamma}$, which is equivalent to (39).

NOTION DE PROXIMITÉ ET ESPACES À STRUCTURE SPHÉROÏDALE.[*]

By Lelio I. Gama.

1. Terminologie adoptée. *Espace accessible* = classe (H) de Fréchet.

Espace régulièrement accessible = espace accessible où à chaque point a correspond une suite évanescente de voisinages, c'est-à-dire une suite de voisinages

$$V_1(a) \supset V_2(a) \supset \cdots \supset V_n(a) \supset \cdots \tag{1}$$

tels que tout $V(a)$ soit un sur-ensemble de $V_n(a)$ pour n suffisamment grand. Les voisinages (1) seront nommés voisinages *principaux* de a; $V_n(a)$ est le voisinage principal de rang n.

Espace strictement accessible = espace accessible où, a et b étant des points distincts, il y a un $V(a)$ et un $V(b)$ tels que $V(a) \cdot V(b) = 0$. C'est, donc, l'espace topologique de Hausdorff.

Nous dirons qu'un espace régulièrement accessible est *uniformément accessible*, si, pour tout voisinage V d'un point a, il en existe un autre $V' \subset V$ et un nombre naturel n' tels que

$$\alpha \in V' \cdot n > n' \rightarrow V_n(\alpha) \subset V.$$

2. Proximité de deux points. Nous dirons que deux points a et b d'un espace régulièrement accessible sont *proches* (l'un de l'autre), s'il existe un nombre naturel n tel que l'on ait

$$a \in V_n(b), \qquad b \in V_n(a).$$

La borne supérieure des n pour lesquels ces conditions se vérifient s'appellera *l'ordre de proximité*, ou, d'une façon abrégée, la *proximité* des points a et b; on la représentera par $\pi(a, b)$. Si a et b sont proches et $a \neq b$, $\pi(a, b)$ est un nombre naturel fini > 0. Si $a = b$, on a $\pi(a, b) = \infty$. Si a et b ne sont pas proches, nous poserons $\pi(a, b) = 0$. L'égalité $\pi(a, b) = n \neq 0$ entraîne, donc, toujours $a \in V_n(b)$, $b \in V_n(a)$.

---

[*] Received March 1, 1944. Several references to the literature suggested by the referee have been added, with the author's consent, by the editors.

3. Application à la théorie générale des suites. Nous dirons que deux suites $\{a_n\}$, $\{b_n\}$ de points d'un espace régulièrement accessible sont *contiguës*, si, étant donné le nombre naturel N, il existe n' tel que $n > n' \rightarrow \pi(a_n, b_n) \geq N$.

Dans un espace uniformément accessible deux suites contiguës ont le même ensemble-limite.[1] *En particulier, si l'une est convergente et a pour limite l, l'autre sera aussi convergente et aura pour limite l.*

Soient $\{a_n\}$ et $\{b_n\}$ contiguës dans un espace uniformément accessible. Posons d'abord $\lim a_n = l$. Prenons un voisinage V de l. Il y a un second voisinage $V' \subset V$ de l et un nombre naturel ν tels que tout voisinage d'ordre ν d'un point situé dans V' soit contenu dans V. Donc on peut déterminer k de façon que les voisinages d'ordre ν des points a_n $(n > k)$ soient tous contenus dans V. En vertu de la contiguïté de $\{a_n\}$ et $\{b_n\}$ on peut déterminer $n' > k$, tel que

$$n > n' \rightarrow \pi(a_n, b_n) \geqq \nu \rightarrow b_n \epsilon V_\nu(a_n).$$

Mais $n > n' \rightarrow V_\nu(a_n) \subset V$, car $n' > k$, donc $n > n' \rightarrow b_n \epsilon V$, et $\lim b_n = b$. Ainsi la convergence d'une des suites vers une limite l entraîne la convergence de l'autre vers la même limite. L'identité des ensembles-limites des deux suites en resultera aisément, car dans un espace régulièrement accessible, l'ensemble-limite d'une suite est l'ensemble des limites de ses sous-suites convergentes.

Nous dirons qu'une suite $\{a_n\}$ de points d'un espace régulièrement accessible est *convergente-C*, si, étant donné le nombre naturel $N > 0$, il existe n' tel que

$$p, q > n' \rightarrow \pi(a_p, a_q) \geq N.$$

On y reconnaît une extension de la condition de Cauchy, ce qui justifie la dénomination adoptée.

Pour que $\{a_n\}$ soit convergente-C il faut et il suffit que deux sous-suites quelconques de $\{a_n\}$ soient contiguës.

Soient

$$a_{\nu(1)}, a_{\nu(2)}, \cdots, a_{\nu(n)}, \cdots$$
$$a_{\mu(1)}, a_{\mu(2)}, \cdots, a_{\mu(n)}, \cdots$$

---

[1] *Point-limite* de $\{a_n\}$ est tout point a tel que, pour tout voisinage V de a, la condition $a_n \epsilon V$ se vérifie pour une infinité de valeurs de l'indice n. L'*ensemble-limite* de $\{a_n\}$ est l'ensemble de ses points-limites. Dans un espace accessible il est toujours fermé, et pour qu'il ne soit pas vide il suffit que le support de la suite (ensemble de ses points distincts) soit compact. Dans un espace régulièrement accessible l'ensemble-limite d'une suite est l'ensemble des limites des sous-suites convergentes.

deux sous-suites (infinies) de la suite convergente-C $\{a_n\}$. (On suppose que les indices ν et μ soient croissants). Le nombre naturel $N > 0$ étant donné, il existe n' tel que

$$(1) \qquad p, q > n' \rightarrow \pi(a_p, a_q) \geqq N.$$

Déterminons n_1 tel que, pour $n > n_1$, on ait à la fois $\nu(n) > n'$, $\mu(n) > n'$. Alors, en vertu de (1),

$$n > n_1 \rightarrow \pi[a_{\nu(n)}, a_{\mu(n)}] \geqq N,$$

ce qui prouve la contiguïté des deux sous-suites considerées. Pour prouver que la condition est suffisante, supposons que $\{a_n\}$ ne soit pas convergente-C, c'est-à-dire qu'il existe un nombre naturel $\nu > 0$ tel que, quel que soit n, on puisse toujours trouver des indices $p, q > n$, $p < q$, qui donnent

$$\pi(a_p, a_q) < \nu.$$

On pourra alors déterminer deux suites (infinies) de nombres naturels croissants, $\{p_n\}$ et $\{q_n\}$, $p_n < q_n$, pour lesquelles $\pi(a_{p_n}, a_{q_n}) < \nu$, et les deux sous-suites correspondantes, $\{a_{p_n}\}$, $\{b_{q_n}\}$ ne seront pas contiguës.

Si l'on convient d'entendre par ensemble *borné* dans un espace régulièrement accessible un ensemble dont tous les points appartiennent à la somme d'un nombre *fini* de voisinages principaux de points de cet espace, on pourra démontrer aisément que

Le support d'une suite convergente-C est un ensemble borné.

Car le nombre naturel $N > 0$ étant choisi arbitrairement, on peut trouver un autre nombre naturel $\nu > 0$ tel que $n > \nu \rightarrow \pi(a_\nu, a_n) \geqq N$, donc

$$n > \nu \rightarrow a_n \in V_N(a_\nu),$$

ce qui montre que tous les points de $\{a_n\}$ sont contenus dans les voisinages d'ordre N des points $a_1, a_2, \cdots, a_\nu$.

On peut même affirmer, comme on le voit, que le support d'une suite convergente-C est *totalement borné* (dans un espace régulièrement accessible), c'est-à-dire qu'il est contenu dans un nombre fini de voisinages d'un rang aussi grand que l'on veut donné d'avance.

Dans un espace uniformément accessible, aucune suite convergente-C n'est discrépante.

En effet, si quelqu'une des sous-suites d'une suite convergente-C a une limite l, il en sera de même, par contiguïté, de toutes les autres, l'espace étant uniformément accessible. Il s'ensuit que l'ensemble-limite de la suite considérée ou bien est vide ou bien n'a qu'un seul élément. (Nous entendons ici par suite *discrépante* une suite dont l'ensemble limite contient au moins deux éléments).

Dans un espace uniformément accessible, toute suite convergente-C, dont le support est compact, est convergente.

Car l'ensemble-limite d'une telle suite se réduit à un seul élément, l, et cela suffit pour que l soit limite de la suite, son support étant compact.

Dans un espace uniformément accessible, la convergence d'une sous-suite d'une suite convergente-C, suffit pour assurer et la convergence de la suite et la compacité de son support.

Soit $\{a_n\}$ la suite convergente-C dont la sous-suite

$$(1) \qquad a_{\nu(1)}, \cdots, a_{\nu(n)}, \cdots$$

a pour limite l. Il suffit de prouver que le support A de $\{a_n\}$ est compact. Supposons, au contraire, que A contienne un sous-ensemble A_1 infini dont le dérivé soit nul. Soit

$$(2) \qquad a_{\mu(1)}, \cdots, a_{\mu(n)}, \cdots$$

la sous-suite de $\{a_n\}$ dont le support est l'ensemble $A_1 - (l)$. Prenons un voisinage V_1 de l tel que

$$V_1[A_1 - (l)] = 0,$$

ce qui est possible car $A'_1 = 0$. Grâce à la condition que l'espace est uniformément accessible, il existe un autre voisinage $V_2 \subset V_1$ de l et un nombre naturel $m > 0$, tels que

$$\alpha \epsilon V_2 \rightarrow V_m(\alpha) \subset V_1,$$

donc

$$\alpha \epsilon V_2 \rightarrow V_m(\alpha)[A_1 - (l)] = 0.$$

Mais on peut déterminer k de façon que $n > k \rightarrow a_{\nu(n)} \epsilon V_2$, donc

$$n > k \rightarrow V_m(a_{\nu(n)})[A_1 - (l)] = 0.$$

Ceci montre que, pour $n > k$, le voisinage d'ordre m du point $a_{\nu(n)}$ de la suite (1) ne contient aucun point de la suite (2), et, par conséquent,

$$n > k \rightarrow \pi[a_{\nu(n)}, a_{\mu(n)}] < m,$$

ce qui est impossible, car $\{a_n\}$ étant convergente-C, les sous-suites (1) et (2) sont contiguës.

De ce qui précède on conclut que

Dans un espace uniformément accessible, la condition nécessaire et suffi-

sante pour qu'une suite convergente-C soit convergente est que son support soit compact.

La notion d'ensemble complet s'étendra aux espaces régulièrement accessibles (= classes (H) de Fréchet à caractère dénombrable). On y dira que l'ensemble E est complet, s'il est vide ou si toute suite convergente-C formée avec des points de E est convergente. En s'appuyant sur les propositions qu'on vient de démontrer, on verra tout de suite que:

Dans un espace uniformément accessible: 1° *tout ensemble compact est complet*; 2° *pour qu'un ensemble E soit complet, il suffit que tout sous-ensemble borné de E soit compact*; 3° *la réunion d'un nombre fini d'ensembles complets est un ensemble complet.*

4. Ensembles clos. Considérons la borne supérieure P des proximités mutuelles des points d'un ensemble E, et appelons *ouverture intérieure* de E le nombre $\omega(E)$ défini par les conditions

$$\omega(E) = 1/P, \text{ si } 0 < P < \infty; \quad \omega(E) = 0, \text{ si } P = \infty; \quad \omega(E) = \infty, \text{ si } P = 0.$$

Cette notion s'applique à tout ensemble situé dans un espace régulièrement accessible. Nous dirons qu'un ensemble E d'un tel espace est *clos*, si $E = 0$, ou si tout sous-ensemble e de E, pour lequel $\omega(e) \neq 0$, est fini. Tout ensemble fini est clos. Dire qu'un ensemble infini E est clos équivaut à dire que, pour tout sous-ensemble infini e de E, on a $\omega(e) = 0$. Dans un espace métrique dire que l'ensemble infini E est clos c'est dire que tout sous-ensemble infini e de E a son diamètre intérieur (= borne inférieure des distances mutuelles des points de e) égal à zéro.

Nous avons déjà convenu de dire qu'un ensemble est totalement borné quand on peut le renfermer dans un nombre fini de voisinages principaux d'un rang aussi élevé que l'on voudra.

Dans un espace régulièrement accessible, tout ensemble clos est totalement borné (la démonstration s'appuie sur l'axiome du choix).

Soit E un ensemble clos bien ordonné. Donnons-nous le nombre naturel $N > 0$. Supposant connu un sous-ensemble $E_n \neq 0$ de E, on peut définir, comme il suit, un sous-ensemble E_{n+1} de E_n: en désignant par a_n le premier élément de E_n, on appellera E_{n+1} l'ensemble des points a de E_n pour lesquels $\pi(a_n, a) < N$. E_{n+1} peut d'ailleurs être vide. En prenant pour E_1 l'ensemble E lui-même, on définit par là une suite d'ensembles décroissants

$$E = E_1 \supset E_2 \supset E_3 \supset \cdots,$$

dont les premiers éléments $a_1, a_2, a_3, \cdots$ vérifient la condition

$$(1) \qquad \pi(a_p, a_q) < N, \qquad (p < q).$$

Nous allons montrer que la suite $a_1, a_2, a_3, \cdots$ est finie. En vertu de (1), les points $a_1, a_2, \cdots$ sont distincts et leurs proximités mutuelles ont une borne supérieure $< N$. L'ensemble A de ces points a donc une ouverture intérieure $\omega(A) \neq 0$. Il en résulte que l'ensemble A, étant un sous-ensemble de l'ensemble clos E, est fini. La suite $\{a_n\}$ se réduit donc bien à un nombre fini de points

$$(2) \qquad a_1, a_2, \cdots, a_k,$$

ce qui revient à supposer $E_{k+1} = 0$. Si alors ξ est un point quelconque de E, il ne peut pas arriver qu'on ait à la fois

$$\pi(a_1, \xi) < N, \ \pi(a_2, \xi) < N, \cdots, \pi(a_k, \xi) < N,$$

puisque le système de ces inégalités entraînerait $\xi \epsilon E_{k+1}$, contrairement à l'hypothèse $E_{k+1} = 0$. Il y a donc au moins un point de la suite (2), soit a_i, pour lequel $\pi(a_i, \xi) \geqq N$, d'où $\xi \epsilon V_N(a_i)$. Ainsi l'on voit que

$$E \subset V_N(a_1) \dotplus V_N(a_2) \dotplus \cdots \dotplus V_N(a_k),$$

c'est-à-dire que E est totalement borné.

5. Espaces à structure sphéroïdale.[2] On peut, moyennant la notion de proximité, définir une classe particulière d'espaces régulièrement accessibles. Nous dirons qu'un tel espace (c'est-à dire une classe (H) de Fréchet à caractère dénombrable) est un espace *à structure sphéroïdale,* si pour chaque nombre naturel $N > 0$, il en existe un autre ν, tel que, a et b étant des points, distincts ou non, de l'espace,

$$(1) \qquad [V_\nu(a) \cdot V_\nu(b) \neq 0] \cdot [\alpha, \beta \,\epsilon\, V_\nu(a) + V_\nu(b)] \rightarrow \pi(\alpha, \beta) \geqq N.$$

Si cette condition se vérifie pour l'indice ν, il en sera évidemment de même lorsqu'on y remplace ν par $\mu > \nu$. La plus petite valeur de l'indice ν pour laquelle (1) se vérifie est le *module de proximité N.*

Il s'ensuit de la définition que, dans un espace à structure sphéroïdale, à tout nombre naturel $N > 0$, il correspond un autre ν tel que, quel que soit le point a de l'espace,

---

[2] A spheroidal space is, by (2), a space of uniform structure in the sense of A. Weil with the added restriction that the uniformity is (by **1** (1)) countable. Cf. A. Weil, "Sur les espaces à structure uniforme et sur la topologie générale," *Actualités scientifiques et industrielles,* no. 551 (1937). Weil's hypothesis U_{III}, p. 7, is essentially 5 (2) of the present paper and his hypotheses U_I and U_{II}, p. 7, are assured by 1 (1).

$$\alpha, \beta \in V_\nu(a) \rightarrow \pi(\alpha, \beta) \geq N. \tag{2}$$

Nous appellerons *sphéroïde de centre a et d'ordre N*, en symboles: $\sigma(a, N)$, le premier voisinage $V_n(a)$ $(n = 1, 2, \cdots)$ pour lequel la condition (2) se trouve vérifiée.

Voici les propriétés fondamentales d'un espace à structure sphéroïdale.

a) $N < N' \rightarrow \sigma(a, N) \supset \sigma(a, N')$.

b) *Quel que soit* N, $\sigma(a, N) \subset V_N(a)$.

[Car $\alpha \in \sigma(a, N) \rightarrow \pi(a, \alpha) \geq N \rightarrow \alpha \in V_N(a)$].

c) *La suite* $\sigma(a, n)$, $(n = 1, 2, \cdots)$, *constitue une famille de voisinages de a équivalente à la famille donnée des voisinages de a.*

d) *Si* ν *est le module de proximité N et si a est un point quelconque de l'espace*

$$n \geq \nu \rightarrow \begin{cases} V_n(a) \subset \sigma(a, N), & \text{(I)} \\ \bar{V}_n(a) \subset V_N(a). & \text{(II)} \end{cases}$$

(I) est une conséquence immédiate des définitions de ν et de $\sigma(a, N)$. Quant à (II), il suffit de montrer que $\bar{V}_\nu(a) \subset V_N(a)$. Soit $\alpha \in \bar{V}_\nu(a)$; tout voisinage de α contient un point de $V_\nu(a)$, donc $V_\nu(a) \cdot V_\nu(\alpha) \neq 0$. Il s'ensuit, compte tenu de ce que ν est le module de proximité N, que $\pi(a, \alpha) \geq N$, d'où $\alpha \in V_N(a)$.

On voit, d'après II, que l'axiome de régularité (cf. Sierpiński, *Intr. to General Topology*, cap. v) est satisfait.

e) *Étant donné le nombre naturel* $N > 0$, *il existe* n' *tel que, pour tout point a de l'espace,*

$$b \in \bar{V}_n(a) \cdot n \geq n' \rightarrow \bar{V}_n(a) \subset V_N(b).$$

Soient ν le module de proximité N et n' le module de proximité ν. Alors, d'après la propriété antérieure, $n \geq n' \rightarrow \bar{V}_n(a) \subset V_\nu(a)$. Soient b un point fixe et β un point variable de $\bar{V}_n(a)$ $(n \geq n')$; b et β appartiennent à $V_\nu(a)$. Par suite, vu que ν est le module de proximité N, on a $\pi(b, \beta) \geq N$, d'où $\beta \in V_N(b)$.

f) *Tout espace à structure sphéroïdale est strictement accessible* [= espace topologique de Hausdorff (à caractère dénombrable)].

Soient, en effet, a et b des points distincts d'un espace à structure sphéroïdale; posons $\pi(a, b) = n_1$ et prenons un nombre naturel $n_2 > n_1$. Soit ν le module de proximité n_2. Les voisinages $V_\nu(a)$ et $V_\nu(b)$ sont disjoints, car $V_\nu(a) \cdot V_\nu(b) \neq 0$ entrainerait $\pi(a, b) \geq n_2 > n_1$, tandis que $\pi(a, b) = n_1$.

g) *Tout espace à structure sphéroïdale est uniformément accessible.*

Soit V un voisinage quelconque de a. Prenons le nombre naturel N assez grand pour que l'on ait $V_N(a) \subseteq V$, et soit ν le module de proximité N. Si $\alpha \in V_\nu(a)$ et $\beta \in V_n(\alpha)$, $n \geq \nu$, on aura $\pi(a, \beta) \geq N$, car $V_\nu(a) \cdot V_n(\alpha) \neq 0$ et ν est le module de proximité N. Donc $\beta \in V_N(a)$. Ainsi

$$\alpha \in V_\nu(a) \cdot n \geq \nu \rightarrow V_n(\alpha) \subset V_N(a) \subseteq V, \qquad \text{c. q. f. d.}$$

6. Métrisation des espaces régulièrement accessibles. Théorème.[3] *Pour qu'un espace régulièrement accessible* [= classe (H) de Fréchet à caractère dénombrable] *soit un espace métrique, il faut et il suffit que ce soit un espace à structure sphéroïdale.*

La condition est nécessaire. Il suffit de prendre pour voisinages principaux des points de l'espace métrique les sphères (ouvertes) de rayon $1/n$ $(n = 1, 2, \cdots)$. On constatera que les limitations[4]

$$\pi(a, b) \geq n, \qquad \rho(a, b) < 1/n$$

sont équivalentes, et, dès lors, si α et β sont des points situés dans $V_{4N}(a) + V_{4N}(b)$, et pourvu que $V_{4N}(a) \cdot V_{4N}(b) \neq 0$, on aura $\rho(\alpha, \beta) < 1/N$, donc $\pi(\alpha, \beta) \geq N$.

La condition est suffisante. En supposant connu le nombre naturel $\nu_n > 0$, posons $\nu_{n+1} = \nu_n + \mu_n$, où l'on désigne par μ_n le module de proximité ν_n dans un espace à structure sphéroïdale donné. En nous donnant arbitrairement la valeur de ν_1, nous aurons par là défini une suite infinie $\{\nu_n\}$ de nombres naturels croissants, où le terme ν_{n+1} est plus grand que le module de proximité ν_n. Appelons F_n le système des voisinages de rang ν_n: F_n forme une couverture de l'espace envisagé. Nous allons montrer, d'après le théorème de Alexandroff et Urysohn (Cf. Fréchet, *Espaces Abstraits*, 1928, p. 220), que $\{F_n\}$ est une suite monotone complète. 1° Soient U'_{n+1}, U''_{n+1} deux éléments, distincts ou non, de F_{n+1} ayant un point p en commun. Soit α un point de U'_{n+1} ou de U''_{n+1}. Comme U'_{n+1} et U''_{n+1} sont des voisinages principaux de rang ν_{n+1} et que ν_{n+1} est plus grand que le module de proximité ν_n, on aura $\pi(p, \alpha) \geq \nu_n$, d'où $\alpha \in V_{\nu_n}(p)$. Par suite U'_{n+1} et U''_{n+1} sont tous les deux contenus dans l'élément $V_{\nu_n}(p)$ de F_n. La suite $\{F_n\}$ est donc bien monotone. 2° Soient ξ un point quelconque de l'espace et $\{U_n\}$ une suite telle que $U_n \in F_n$, $\xi \in U_n$. Il s'agit de montrer que $\{U_n\}$ constitue une famille de voisinages de ξ équivalente à la famille donnée des voisinages $V(\xi)$ de ξ. Que tout U_n contient un $V(\xi)$ découle immédiatement de ce qu'on a affaire à une classe (H) de Fréchet. Réciproquement, soit $V(\xi)$ un voisinage donné de ξ. Prenons N de façon que

---

[3] Cf. A. Weil, *loc. cit.*[1] and J. W. Tukey, *Convergence and Uniformity in Topology*, Princeton (1940), p. 61, Theorem 6. 1.

[4] $\rho(a, b)$ = distance des points a et b.

$V_N(\xi) \subset V(\xi)$. Par la propriété (*e*) des espaces à structure sphéroïdale (4), il existe n' tel que, quel que soit le point a de l'espace,

$$\xi \epsilon V_n(a) \cdot n \geq n' \rightarrow V_n(\alpha) \subset V_N(\xi) \subset V(\xi).$$

Comme U_n est un voisinage de rang ν_n d'un certain point a de l'espace, il suffit de prendre $\nu_n \geq n'$ pour que l'inclusion $U_n \subset V(\xi)$ se trouve vérifiée.

7. Rôle de la notion de proximité dans les espaces métriques. D'après le théorème qui précède, il est intéressant d'examiner le rôle que joue la notion de proximité dans les démonstrations des propriétés des espaces métriques, en cherchant à retrouver ces propriétés directement dans un espace à structure sphéroïdale. On y trouvera, peut-être, une méthode assez simple d'étudier les propriétés des espaces métriques sans faire appel à la notion de distance. Nous nous bornerons, dans ce numéro, à deux théorèmes classiques.

I. *La condition de normalité* (au sens strict). *Si A et B sont des ensembles séparés ($A\bar{B} + \bar{A}B = 0$), on peut déterminer deux ensembles ouverts P, Q, tels que $A \subset P$, $B \subset Q$, $PQ = 0$.*

À chaque point a de A associons le premier voisinage principal, $V_M(a)$, tel que $V_M(a) \cdot B = 0$, et, en appelant μ le module de proximité M, posons

$$P = \Sigma V_\mu(a),$$

la sommation s'étendant à tous les points a de A (μ varie, en général, avec le point a considéré). De même, associons à chaque point b de B le premier voisinage principal, $V_N(b)$, tel que $V_N(b) \cdot A = 0$, et, désignant par ν le module de proximité N, posons

$$Q = \Sigma V_\nu(b).$$

Il suffit de montrer que $PQ = 0$. Supposons, au contraire, que $k \epsilon PQ$. Le point k appartiendra à un couple d'éléments $V_{\mu'}(a') \epsilon P$ et $V_{\nu'}(b') \epsilon Q$, et, si $V_{M'}(a')$, $V_{N'}(b')$ sont les voisinages associés aux points a', b', les indices μ', ν' seront, respectivement, les modules de proximité M' et de proximité N' Soit $\nu' \geq \mu'$. Alors, comme il s'agit d'un espace à structure sphéroïdale, on aura, pour deux points quelconques α, β pris dans $V_{\mu'}(a') + V_{\nu'}(b')$, $\pi(\alpha, \beta) \geq M'$. On a, en particulier, $\pi(a', b') \geq M'$, et, par suite, $b' \epsilon V_{M'}(a')$, ce qui est impossible, d'après la définition même de $V_{M'}(a')$.

II. *Le théorème de Borel-Lebesgue.* Pour démontrer ce théorème dans un espace à structure sphéroïdale, on peut employer le mode de raisonnement utilisé par M. Fréchet dans sa démonstration directe du théorème pour le cas des espaces métriques,[5] en s'appuyant aussi sur l'axiome du choix. Soit K l'ensemble compact et fermé, couvert par la famille infinie ΣA, de sorte que $K \subset \Sigma I$, I étant l'intérieur de A. Nous considérons K comme un ensemble bien ordonné. Associons à chaque point a de K le premier voisinage principal de a, $U(a)$, ayant la propriété d'être contenu dans un I. Il suffit de prouver que K peut être couvert par une collection finie des voisinages $U(a)$. Supposons que cela n'ait pas lieu. On peut alors, en partant du premier élément a_1 de K, et en utilisant le bon ordre de K, définir une suite infinie $\{a_n\}$ de points de K, de telle façon que a_n $(n > 1)$ ne soit contenu dans aucun des voisinages $U(a_1), \cdots, U(a_{n-1})$, associés, comme on l'a expliqué, aux points $a_1, \cdots, a_{n-1}$. Les a_n sont évidemment des points distincts. L'ensemble A des a_n étant compact (puisque $A \subset K$), on peut extraire de $\{a_n\}$ une sous-suite $\{\alpha_n\}$ convergente vers un point ξ du dérivé A'; ce point ξ appartient d'ailleurs à K (puisque K est fermé). Envisageons le voisinage $U(\xi)$, associé au point ξ. En vertu de l'uniforme accessibilité d'un espace à structure sphéroïdale, il existe un voisinage de ξ, $W \subset U(\xi)$, et un nombre naturel n', tels que

$$\alpha \in W \cdot n > n' \rightarrow V_n(\alpha) \subset U(\xi).$$

Soit donc $N > n'$ un nombre naturel assez grand pour que l'on ait $\sigma(\xi, N) \subset W$, et considérons deux points α_i, α_{i+1} de la suite $\{\alpha_n\}$ contenus dans $\sigma(\xi, N)$, ce qui est possible, car $\lim \alpha_n = \xi$. On aura $\pi(\alpha_i, \alpha_{i+1}) \geqq N$, d'où

$$\alpha_{i+1} \in V_N(\alpha_i). \tag{1}$$

Mais, d'autre part, puisque $N > n'$, $V_N(\alpha_i) \subset U(\xi)$, et, par suite, $V_N(\alpha_i)$ est contenu dans un I; il en résulte que $V_N(\alpha_i) \subset U(\alpha_i)$, puisque $U(\alpha_i)$ est le premier voisinage principal de α_i ayant la propriété d'être contenu dans un I. Comme, d'après la définition même de la suite $\{a_n\}$, α_{i+1} n'est pas contenu dans $U(\alpha_i)$, on aura

$$\alpha_{i+1} \bar{\epsilon} V_N(\alpha_i). \tag{2}$$

Les conclusions (1) et (2) sont contradictoires.

III. *Remarque sur les ensembles totalement bornés.* Il nous sera utile d'avoir observé que, si E est un ensemble totalement borné dans un espace à structure sphéroïdale, il existe, pour chaque naturel $n > 0$, une collection finie de points de E, $\xi_1, \cdots, \xi_k$, tels que

$$E \subset V_n(\xi_1) \dotplus V_n(\xi_2) \dotplus \cdots \dotplus V_n(\xi_k).$$

---

[5] *American Journal of Mathematics*, vol. 50 (1928), p. 70.

Étant donné N, on peut, en effet, déterminer ν tel que, pour tout point a de l'espace et pour tout point $b \in \bar{V}_\nu(a)$, on ait (**5**, e)

$$(1) \qquad \bar{V}_\nu(a) \subset V_N(b).$$

D'autre part, puisque l'ensemble E est totalement borné, il admet une inclusion de la forme

$$E \subset V_\nu(a_1) \dotplus V_\nu(a_2) \dotplus \cdots \dotplus V_\nu(a_k),$$

avec $EV_\nu(a_s) \neq 0$ $(s = 1, 2, \cdots, k)$. Alors, si $\xi_s \in EV_\nu(a_s)$, on aura, d'après (1),

$$V_\nu(a_s) \subset V_N(\xi_s), \qquad (s = 1, 2, \cdots, k)$$

donc

$$E \subset V_N(\xi_1) \dotplus V_N(\xi_2) \dotplus \cdots \dotplus V_N(\xi_k).$$

8. Suites évanescentes d'ensembles.[6] Nous allons étudier les propriétés des suites évanescentes d'ensembles, définies dans un espace métrisable, en utilisant encore la notion de proximité au lieu de celle de distance, c'est-à-dire en prenant pour base un espace à structure sphéroïdale. Nous appellerons *suite évanescente d'ensembles* (dans un espace régulièrement accessible) toute suite $\{E_n\}$ d'ensembles satisfaisant aux deux conditions suivantes:

1° quels que soient les indices p, q, $\bar{E}_p \bar{E}_q \neq 0$;

2° à chaque E_n correspond un point a_n de $\bar{E}_n$ ayant cette propriété: étant donné le nombre naturel $\nu > 0$, il existe n' tel que $n > n' \rightarrow E_n \subset \bar{V}_\nu(a_n)$.

Le point a_n sera nommé le *centre* de E_n. Si une des suites $\{E_n\}$, $\{\bar{E}_n\}$ est évanescente, il en est de même de l'autre.

a) *Dans un espace à structure sphéroïdale, si* $\{E_n\}$ *est une suite évanescente et si* a_n *est le centre de* E_n, *alors étant donné le nombre naturel* $\nu > 0$, *il existe* n' *tel que* $n > n' \rightarrow \bar{E}_n \subset \sigma(a_n, \nu)$, *donc aussi* (**5**, b) $\bar{E}_n \subset V_\nu(a_n)$.

Grâce à la propriété d) (n. 5), on peut déterminer m de façon qu'on ait, quel que soit le point a de l'espace, $V_m(a) \subset \sigma(a, \nu)$, et, en suite, déterminer μ de façon que $\bar{V}_\mu(a) \subset V_m(a)$. On aura alors, quel que soit a, $\bar{V}_\mu(a) \subset \sigma(a, \nu)$. Or, d'après la définition d'une suite évanescente, on a $\bar{E}_n \subset \bar{V}_\mu(a_n)$ dès que l'indice n soit suffisamment grand.

b) *Si* $\{E_n\}$ *est évanescente dans un espace à structure sphéroïdale,* 1° *toute suite* $\{\xi_n\}$, *où* $\xi_n \in \bar{E}_n$, *est convergente-C*; 2° *si l'on prend dans chaque* $\bar{E}_n$ *un couple de points* ξ_n, η_n, *les suites* $\{\xi_n\}$, $\{\eta_n\}$ *seront contiguës.*

---

[6] Cf. G. Birkhoff, "Moore-Smith convergence in general topology," *Annals of Mathematics*, vol. 38 (1937), pp. 39-60.

limite ω (c'est ce qui résulte des théorèmes d) et e) du n. précédent). Nous allons montrer que

Dans un espace à structure sphéroïdale, il y a identité entre les notions d'ensemble complet et d'ensemble cantorien.

Soit $E \neq 0$ un ensemble complet. Considérons une suite évanescente $\{V(a_n)\}$, dont chaque terme est un voisinage principal d'un point a_n $(n = 1, 2, \cdots)$ de E. $\{a_n\}$ est une suite convergente-C (**8**, b), qui a une limite ω, puisque E est complet. Donc (**8**, c) la suite $\{V(a_n)\}$ a pour point-limite ω, et, par suite, E est un ensemble cantorien.

Soit $E \neq 0$ un ensemble cantorien. Considérons une suite convergente-C, $\{a_n\}$, dont les termes sont des points de E. Soient $n(1)$ le plus petit nombre naturel pour lequel

$$n > n(1) \rightarrow \pi[a_{n(1)}, a_n] \geq 1;$$

$n(2)$ le plus petit nombre naturel $> n(1)$, pour lequel

$$n > n(2) \rightarrow \pi[a_{n(2)}, a_n] \geq 2,$$

et ainsi de suite. Il est manifeste qu'on définit par cette norme une suite (infinie) de nombres naturels croissants $n(1), n(2), \cdots, n(\nu), \cdots$ ayant cette propriété:

$$(1) \qquad n > n(\nu) \rightarrow \pi[a_{n(\nu)}, a_n] \geq \nu.$$

Envisageons la suite suivante de voisinages principaux des points $a_{n(\nu)}$:

$$(2) \qquad V_1[a_{n(1)}].\ V_2[a_{n(2)}], \cdots, V_\nu[a_{n(\nu)}], \cdots.$$

Si μ est un nombre naturel $> \nu$, on a $n(\mu) > n(\nu)$, et, d'après (1), $\pi[a_{n(\nu)}, a_{n(\mu)}] \geq \nu$, donc

$$\mu > \nu \rightarrow a_{n(\mu)} \in V_\nu[a_{n(\nu)}].$$

Ceci montre que chaque voisinage de la suite (2) contient les « centres » $a_{n(\nu)}$ de tous les voisinages suivants, et, par conséquent, cette suite remplit la condition 1° d'évanescence (**8**); elle satisfait en outre à la condition 2°, car elle est formée de voisinages principaux dont les rangs ν sont croissants. Et alors, puisque E est un ensemble cantorien, la suite (2) possède un point-limite ω, et on aura

$$\lim_{\nu=\infty} a_{n(\nu)} = \omega.$$

Mais $\{a_{n(\nu)}\}$, $\nu = 1, 2, \cdots$, est une sous-suite de la suite convergente-C $\{a_n\}$; il s'ensuit (**3**) que $\omega = \lim a_n$. Donc E est un ensemble complet.

10. Propriétés transitives. Nous dirons qu'une propriété P d'ensemble est *transitive,* si, pour tout ensemble non vide E jouissant de cette propriété, la condition suivante est remplie:

a) Pour toute décomposition $E = E_1 \dotplus E_2$, l'une au moins des parties E_1, E_2 possède la propriété P.

Si l'on ajoute à cette condition la restriction suivante:

b) si le sous-ensemble e_1 de E ne jouit pas de la propriété P, il en sera de même de tout sous-ensemble (non vide) de e_1,

alors la propriété P deviendra ce que nous avons appelé une propriété *héréditaire.*[7]

Si $E \subset E_1 \dotplus E_2 \dotplus \cdots \dotplus E_m$ $(E \neq 0)$, toute propriété transitive ou héréditaire de E se transmet à l'un au moins des produits $EE_1, EE_2, \cdots, EE_m$ qui ne soient pas vides.

Nous dirons qu'une propriété P d'un ensemble E $(\neq 0)$ est *localisable* en un point a, si tout voisinage de a contient un sous-ensemble $e \neq 0$ de E possédant la propriété P. [Cette définition de la localisation d'une propriété est un peu moins restrictive que celle que nous avons adopté dans l'article cité].

Dans un espace à structure sphéroïdale, toute propriété transitive d'un ensemble totalement borné et complet E $(\neq 0)$ est localisable en un point au moins de la fermeture de E.[8]

[Avec la définition plus restrictive de la localisation d'une propriété, ce théorème est valable pour les propriétés héréditaires. C'est là le point de vue où nous nous sommes placés dans l'article déjà cité].

Pour le démontrer, soit $E \neq 0$ un ensemble totalement borné et complet (dans un espace à structure sphéroïdale) jouissant d'une propriété transitive P. Nous considérons E comme un ensemble bien-ordonné. E étant totalement borné, il existe, pour chaque nombre naturel $n > 0$, des collections finies formées de voisinages de rang n de points appartenant à E, chaque collection couvrant E. D'après le principe du choix, on peut donc, à chaque nombre naturel $n > 0$, faire correspondre une telle collection, soit K_n. Les éléments de K_n (voisinages de rang n de points de E) seront censés ordonnés dans l'ordre même de ces points.

---

[7] "Sur quelques points de la théorie des ensembles abstraits et la notion d'accumulatif," *Anais da Academia Brasileira de Ciencias,* vol. 12 (1940). Dans la définition donnée dans cet article nous avions employé l'expression «l'ensemble E» dans un sens générique, en y laissant donc sous-entendu que les deux conditions *a*) et *b*) devraient être remplies pour *tout* ensemble E ayant la propriété envisagée.

[8] La démonstration s'appuie sur l'axiome du choix.

Puisque $E \subset K_1$, il existe un premier voisinage U_1 dans la collection K_1, tel que

$$EU_1 \neq 0, \qquad EU_1 \in (P),$$

(P) désignant la classe des ensembles qui jouissent de la propriété considérée P. On peut alors définir une suite infinie $U_1, \cdots, U_n, \cdots, (U_n \in K_n)$, ayant cette propriété: quel que soit n,

$$(1) \qquad EU_1U_2 \cdots U_n \neq 0,$$

$$(2) \qquad EU_1U_2 \cdots U_n \in (P).$$

Car, une fois determinée une suite finie $U_1, \cdots, U_m$ remplissant les conditions (1), (2) pour $n = 1, 2, \cdots, m$, le produit $EU_1 \cdots U_m$, qui est un sous-ensemble non vide de E, est contenu dans K_{m+1}, et, puisqu'il a la propriété P, on trouvera dans la collection K_{m+1} un premier voisinage, U_{m+1}, tel que

$$0 \neq EU_1 \cdots U_m U_{m+1} \in (P).$$

Il est maintenant aisé de voir que la suite $\{U_n\}$ ainsi définie est une suite évanescente (**8**); car, d'abord, p et q étant des nombres naturels > 0, on a, en vertu de la condition (1), $U_pU_q \neq 0$, ce qui montre que la condition 1° d'évanescence est remplie. D'autre part, soit $\nu > 0$ un nombre naturel, et désignons par a_n le point de E dont U_n est un voisinage. Il est clair que

$$n > \nu \rightarrow U_n(a_n) \subset V_\nu(a_n) \subset \bar{V}_\nu(a_n),$$

ce qui fait voir que la condition 2° d'évanescence est aussi remplie.

E est un ensemble cantorien, parce qu'il est complet dans un espace à structure sphéroïdale (**9**). Il s'ensuit que $\{U_n\}$ admet un point-limite ω. Soit V un voisinage quelconque de ω: il y a un $U_m \subset V$, et, par suite, $EU_1 \cdots U_m \subset V$. Mais l'ensemble $EU_1 \cdots U_m$ possède la propriété P d'après la condition (2). La propriété P est donc bien localisable au point ω. L'on voit en même temps que $\omega \in \bar{E}$, comme on devait s'y attendre. Le théorème est donc démontré.

En voici une application.

Dans un espace à structure sphéroïdale, la condition nécessaire et suffisante pour qu'un ensemble soit compact, est que cet ensemble soit totalement borné et complet.

Soit, en effet, E un sous-ensemble infini d'un ensemble totalement borné et complet (dans un espace à structure sphéroïdale). E sera lui-même un ensemble totalement borné et complet. Or, la propriété pour un ensemble d'être infini est transitive. Il y a donc un point ω dont tout voisinage contient une infinité de points de E. Mais alors $\omega \in E'$, donc $E' \neq 0$. La condition est donc suffisante. La condition est nécessaire, parce que, d'abord, dans un espace uniformément accessible quelconque, un ensemble compact est toujours complet, et, d'autre part, il est facile de voir que, dans un espace à structure sphéroïdale, un ensemble compact est toujours clos (et, par suite, totalement borné, **4**). En effet, si $a \in E'$, et si $N > 0$ est un nombre naturel quelconque, il y a un point $b \neq a$ de E contenu dans le sphéroïde $\sigma(a, N)$ (**5**, c); par suite $\pi(a, b) \geq N$ et $\omega(E) = 0$ (**4**).

11. Note sur le théorème de Borel-Lebesgue. D'après cela, le théorème du numéro précédent sur la localisation des propriétés transitives peut être énoncé sous cette forme:

Dans un espace à structure sphéroïdale, toute propriété transitive d'un ensemble compact E ($\neq 0$) est localisable en un point au moins de la fermeture de E.

Sous cette forme, cette proposition peut être établie bien aisément en s'appuyant sur le théorème de Borel-Lebesgue. Mais il est intéressant de remarquer[9] que ce dernier théorème n'est lui-même qu'une forme particulière de la proposition qu'on vient d'énoncer, c'est-à-dire que la propriété de Borel-Lebesgue découle, dans les espaces à structure sphéroïdale, d'un processus de localisation d'une propriété transitive (ou héréditaire, lorsqu'on adopte la définition plus restrictive du concept «localisable»).[9] Voici comment on peut le montrer.

Soit E un ensemble compact et fermé couvert par une famille infinie F d'ensembles ϕ, dans un espace à structure sphéroïdale. Il est manifeste que la propriété P pour un ensemble de n'être pas couvert par une collection finie des ensembles ϕ est transitive. Supposons, alors, que l'ensemble donné E ait cette propriété P. Soit $\omega \in \bar{E} = E$ un point de localisation de la propriété P. Le point ω (parce qu'il appartient à E) est dans l'intérieur I d'un ensemble ϕ_1 de la famille de couverture F. Soit V un voisinage de ω tel que $V \subset I$. Puisque la propriété P est localisable au point ω, V contient un sous-ensemble e de E qui n'est couvert par aucune collection finie des ensembles ϕ, ce qui est en contradiction avec l'inclusion $e \subset \phi_1$ ($e \subset V \subset I \subset \phi_1$).

Observatorio Nacional,
Rio de Janeiro, Brazil.

---

[9] Voir l'article déjà cité *Sur quelques points* etc.

DUALITY THEOREMS FOR GENERALIZED MANIFOLDS.*

By Edward G. Begle.[1]

In a previous paper[2] we gave a new definition of a generalized manifold.[3] By combining cohomology theory and some of our results on homology local connectedness, we were able to give a short proof of the Poincaré Duality Theorem for these manifolds. The object of the present paper is to show that the same methods can be used to give equally simple proofs of several other duality theorems, including, in particular, the Alexander Duality Theorem, for generalized manifolds.

The proofs require a discussion of open generalized manifolds, i. e., locally compact spaces with the local homology and cohomology properties of Euclidean spaces. Accordingly, the first four sections below are devoted to the required definitions and the changes necessitated in passing from the compact to the locally compact case. The Poincaré Duality Theorem is proved in Section **5** and is used in Section **6** in proving other duality theorems. We conclude in Section **7** by showing that the algebraic isomorphisms established in the Alexander Duality Theorem also have a useful geometric property.

1. Homology and cohomology theory of locally compact spaces. In addition to the ordinary Čech homology and cohomology groups of a topological space, there are two other pairs of groups which are necessary for the treatment of open generalized manifolds. In the following, the coefficient group is assumed to be an arbitrary but fixed field.

Let $\mathfrak{S}$ be a locally compact[4] space and let $\{L_\lambda\}$ be the collection of all compact subsets of $\mathfrak{S}$. We partially order the collection $\{L_\lambda\}$ by: $L_\lambda < L_\mu$ if, and only if, $L_\lambda \supset L_\mu$. Since the sum of two compact sets is compact, it is

---

* Received March 13, 1943.

[1] Most of the results of this paper were obtained while the author was a National Research Fellow.

[2] "Locally connected spaces and generalized manifolds," *American Journal of Mathematics*, vol. 64 (1942), pp. 553-574. This paper will hereafter be referred to as [B]. Numbers in square brackets refer to the bibliography at the end of [B]. A knowledge of [B] is necessary for the understanding of the present paper.

[3] Our generalized manifold includes and is very similar to Čech's [5]. If the space in question is compact metric, both types are the same as Wilder's [20].

[4] Compact = bicompact. All spaces considered here are understood to be normal.

clear that if L_μ and L_ν are arbitrary then $L_\lambda = L_\mu \cup L_\nu$ is such that $L_\lambda < L_\mu$. $L_\lambda < L_\nu$, i. e., $\{L_\lambda\}$ is a directed set [17].

Let z_λ^p denote the group of p-dimensional Čech cycles of $\mathfrak{S}$ which are carried[5] by L_λ, and let f_λ^p denote the subgroup of z_λ^p consisting of those cycles which bound on L_λ. If $L_\lambda < L_\mu$, the identity mapping is a homomorphism $\sigma^{*}{}_\lambda^\mu$ of z_μ^p into z_λ^p. Hence $\{z_\lambda^p, \sigma^{*}{}_\lambda^\mu\}$ and $\{f_\lambda^p, \sigma^{*}{}_\lambda^\mu\}$ are direct systems of groups [see 15, p. 57]. We denote their limit groups by $z^p(\mathfrak{S})$ and $f^p(\mathfrak{S})$. The quotient group, $h^p(\mathfrak{S})$, is called the p-dimensional compact homology group of $\mathfrak{S}$. The elements of $z^p(\mathfrak{S})$ are called compact cycles.

Note that since $\{z_\lambda^p, \sigma^{*}{}_\lambda^\mu\}$ is a direct system, a compact cycle of $\mathfrak{S}$ is determined, up to homology, by any one of its coordinates.

Next let Z_p^λ be the group of p-dimensional Čech cocycles of $\mathfrak{S}$ mod $\mathfrak{S} - L_\lambda$ and let F_p^λ be the subgroup of those cocycles which are ~ 0 mod $\mathfrak{S} - L_\lambda$. The identity mapping is a homomorphism σ_λ^μ of Z_p^λ into Z_p^μ. Hence $\{Z_p^\lambda, \sigma_\lambda^\mu\}$ and $\{F_p^\lambda, \sigma_\lambda^\mu\}$ are inverse systems of groups. The limit groups are denoted by $Z_p(\mathfrak{S})$ and $F_p(\mathfrak{S})$ and their quotient group by $H_p(\mathfrak{S})$. The latter is called the p-dimensional infinite cohomology group of $\mathfrak{S}$ and the elements of $Z_p(\mathfrak{S})$ are called infinite cocycles.

If Υ^p is a compact cycle and Γ_q an infinite cocycle, $q \leqq p$, the collection $\{\Upsilon_\lambda^p \cdot \Gamma_q^\lambda\}$ is easily seen to be a compact cycle, Υ^{n-p}, and we define this to be the intersection, $(\Upsilon^p \cdot \Gamma_q)$ of Υ^p and Γ_q.

The Kronecker Index, $KI(\Upsilon^p, \Gamma_p)$, induces a pairing of $h^p(\mathfrak{S})$ and $H_p(\mathfrak{S})$ to the coefficient field, and, as Lefschetz has shown, these two groups are orthogonal under this pairing. Hence [15, pp. 262, 77]:

THEOREM L_1. *If either one of the groups $h^p(\mathfrak{S})$, $H_p(\mathfrak{S})$ has a finite basis, there is an isomorphism θ_1 of $h^p(\mathfrak{S})$ onto $H_p(\mathfrak{S})$ such that if Υ^p is any basis element of $h^p(\mathfrak{S})$, then $KI(\Upsilon^p \cdot \theta_1(\Upsilon^p)) = 1$.*

Paralleling these groups are the groups of compact cocycles and infinite cycles of $\mathfrak{S}$, denoted by $h_p(\mathfrak{S})$ and $H^p(\mathfrak{S})$. These are defined in the same manner as the above groups, but with the family $\{L_\lambda\}$ replaced by the family $\{O_\lambda\}$ of all open subsets of $\mathfrak{S}$ with compact closures. For the details see [15, p. 262]. In this case a compact cocycle is a direct system and is defined up to cohomology by any one of its coordinates. Here also the intersection of Γ^p and Υ_q, $q \leqq p$, can be defined and is a compact cycle of dimension $p - q$.

---

[5] A set L carries a cycle Γ if all coordinates of Γ, as well as all chains used in establishing the homologies between different coordinates, are on L (in the sense of [B, p. 561]). L is called a carrier of Γ. A similar definition holds for cocycles, the carrier this time being an open set.

It is clear that if $\mathfrak{S}$ is compact, then $h^p(\mathfrak{S}) \simeq H^p(\mathfrak{S})$ and $h_p(\mathfrak{S}) \simeq H_p(\mathfrak{S})$.

THEOREM L_2. *If either one of the groups* $H^p(\mathfrak{S})$, $h_p(\mathfrak{S})$ *has a finite basis, there is an isomorphism* θ_2 *of* $H^p(\mathfrak{S})$ *onto* $h_p(\mathfrak{S})$ *such that if* Γ^p *is a basis element of* $H^p(\mathfrak{S})$, *then* $KI(\Gamma^p \cdot \theta_2(\Gamma^p)) = 1$.

Since these isomorphisms do not exist in the case of infinite bases, our results below hold only for groups with finite bases. It is therefore understood without further mention that all the homology groups below are assumed to have finite bases, and all Betti numbers and local Betti numbers are assumed to be finite.

In many cases the locally compact space $\mathfrak{S}$ appears as an open subset of another locally compact space $\mathfrak{T}$, and it is desirable to relate the groups of $\mathfrak{S}$ with those of $\mathfrak{T}$. Denote by $H^p(\mathfrak{T} \bmod (\mathfrak{T} - \mathfrak{S}))$ the group of p-cycles of $\mathfrak{T} \bmod (\mathfrak{T} - \mathfrak{S})$ reduced modulo those which are $\sim 0 \bmod (\mathfrak{T} - \mathfrak{S})$. Similarly, let $H_p(\mathfrak{T} \bmod (\mathfrak{T} - \mathfrak{S}))$ be the group of p-cocycles of $\mathfrak{T}$ which are in $\mathfrak{S}$ reduced modulo those which are ~ 0 in $\mathfrak{S}$.

THEOREM 1.1. *If the closure of* $\mathfrak{S}$ *in* $\mathfrak{T}$ *is compact, then* $H_p(\mathfrak{T} \bmod (\mathfrak{T} - \mathfrak{S}))$ *is isomorphic to* $h_p(\mathfrak{S})$ *and* $H^p(\mathfrak{T} \bmod (\mathfrak{T} - \mathfrak{S}))$ *is isomorphic to* $H^p(\mathfrak{S})$.

Proof. The second part of this theorem follows from Theorem L_2 and from Theorem (4.7) of Chapter VII of [15]. Hence we need only prove the first isomorphism.

Let then Γ_p be a cocycle of $\mathfrak{T}$ which lies in $\mathfrak{S}$ and let $\mathfrak{U}$ be a covering (all coverings are understood to be finite) of $\mathfrak{T}$ such that Γ_p has a coordinate $\Gamma_p{}^u$ on U. By a theorem due to A. D. Wallace [15, p. 263], there is a refinement $\mathfrak{V}$ of $\mathfrak{U}$ and a neighborhood O of $\mathfrak{T} - \mathfrak{S}$ such that $\pi^*{}_v{}^u\Gamma_p{}^u$ is on $\mathfrak{T} - O$, a compact subset of $\mathfrak{S}$. In the same way, if Υ_p is a compact cocycle of $\mathfrak{S}$ such that $\Upsilon_p \sim 0$ in $\mathfrak{S}$, then $\pi^*{}_v{}^u\Upsilon_p{}^u \sim O$ in a compact subset of $\mathfrak{S}$. This proves the theorem.

We recall that in definitions of local connectedness a 0-chain is to be considered a 0-cycle only if the sum of its coefficients is zero. In defining local co-Betti numbers, it is necessary to make an analogous convention regarding 0-cocycles. The most convenient method of doing this is to augment[6] all complexes involved, i. e. to add to each complex a new element of dimension -1 which has with each 0-simplex the incidence number 1. The homology and cohomology groups based on augmented complexes are denoted, by H_a^p, H_p^a etc. Clearly, for $p > 0$, H^p and H_a^p are the same. However, H_a^0 has one less generator than H^0, and similarly for H_0^a and H_0. The duality theorems above still hold in the augmented case.

---

[6] For a complete discussion of this notion, see [15, p. 130].

In the following, in all definitions of local connectedness, local Betti numbers, and local co-Betti numbers, all the complexes involved are understood to be augmented. In all other cases all complexes are unaugmented unless the contrary is explicitly stated.

2. Local connectedness in locally compact spaces.

For such spaces, Čech's definition of local connectedness [8] is most useful.

Definition 2.1. $\mathfrak{S}$ is $q - lc$ at a point s if given a neighborhood P of s and a covering $\mathfrak{B}$ of $\mathfrak{S}$, there exists a neighborhood $Q = Q^q(P) \subset P$ and a covering $\mathfrak{A} = \mathfrak{A}^q(\mathfrak{B}, P) < \mathfrak{B}$ such that if X^q is a cycle on $A \cap Q$, then $X^q = FY^{q+1}$, where Y^{q+1} is a chain of $(A \cup B) \cap P$. $\mathfrak{S}$ is $q - lc$ if it is $q - lc$ at every point, and $\mathfrak{S}$ is lc^p if it is $q - lc$ for all $q \leq p$.

Definition 4.2 of [B] is a definition of uniform local connectedness, and hereafter the notations $q - ulc$, ulc^p, etc., will be used for this property. If $\mathfrak{S}$ is compact, then definitions 4.2 of [B] and 2.1 are equivalent, but in general this is not the case. In particular, Theorem 5.5 of [B] is not a consequence of Definition 2.1. However, we can prove an analogous but weaker theorem which is sufficient for our purposes.

THEOREM 2.1. *Let L be a compact subset of $\mathfrak{S}$ and O a neighborhood of L. If $\mathfrak{S}$ is lc^p, then for each pair $\mathfrak{E}$ and $\mathfrak{B}$ of coverings of $\mathfrak{S}$, there exist coverings $\mathfrak{H} = \bar{\mathfrak{H}}^p(\mathfrak{E}; L, O)$ and $\mathfrak{A} = \bar{\mathfrak{A}}^p(\mathfrak{B}, \mathfrak{E}; L, O)$ such that if K is a complex of dimension $\leqq p + 1$ and τ' is a partial realization of K on $A \cap L$ of norm $< \mathfrak{H}$, then τ' can be extended to a realization τ of K on $(A \cup B) \cap O$ of norm $\mathfrak{E}$.*

Proof. For each $q \leq p$ and each compact subset L of S, it follows, by the usual argument, that for each covering $\mathfrak{E}$ of $\mathfrak{S}$ there is a refinement $\mathfrak{H}^q(\mathfrak{E}; L)$ and for each covering $\mathfrak{B}$ there is a refinement $\mathfrak{A}^q(\mathfrak{B}, \mathfrak{E}; L)$ such that if X^q is a cycle of $\mathfrak{A} \cap L$ of diameter $< \mathfrak{H}^q(\mathfrak{E}, L)$, then $X^q = FY^{q+1}$ where Y^{q+1} is a chain on $A \cup B$ of diameter $< \mathfrak{E}$.

Now suppose that the theorem has been proved for $q - 1$. Choose an open set O_1 such that $L \subset O_1 \subset \bar{O}_1 \subset O$ and such that $\bar{O}_1$ is compact. Given $\mathfrak{E}$, choose $\mathfrak{E}_1 < \mathfrak{E}$ such that $St(\bar{O}_1, \mathfrak{E}_1) \subset O$. Then choose $\mathfrak{E}_2 \mathbin{{}^*{<}} \mathfrak{H}^q(\mathfrak{E}_1, \bar{O}_1)$ and set $\bar{\mathfrak{H}}^q(\mathfrak{E}; L, O) = \bar{\mathfrak{H}}^{q-1}(\mathfrak{E}_2; L, O_1)$. Given $\mathfrak{B}$, let $\mathfrak{B}_1 = \mathfrak{A}^q(\mathfrak{B}, \mathfrak{E}_1; \bar{O}_1)$ and set $\bar{\mathfrak{A}}^q(\mathfrak{B}, \mathfrak{E}; L, O) = \bar{\mathfrak{A}}^{q-1}(\mathfrak{B}_1, \mathfrak{E}_2; L, O_1)$.

That these choices of $\mathfrak{H}$ and $\mathfrak{A}$ are satisfactory follows by essentially the same argument as used in the proof of Theorem 1. 1 of [B].

The proof of Theorem 5. 5 of [B], using Theorem 2. 1 in the appropriate places, now yields

THEOREM 2. 2. *Let* $\mathfrak{S}$ *be* lc^p, *L a compact subset of* $\mathfrak{S}$, *and O a neighborhood of L. Then there is a covering* $\mathfrak{U}_0$ *of* $\mathfrak{S}$ *such that if* Γ^q *and* $\bar{\Gamma}^q$, $q \leq p$, *are cycles of* $\mathfrak{S}$ *with* $\Gamma_0^q \sim \bar{\Gamma}_0^q$ *on* $U_0 \cap L$, *then* $\Gamma^q \sim \bar{\Gamma}^q$ *in O.*

The following corollary, which is an immediate consequence of the above theorem, generalizes a theorem due to Wilder [21] from locally compact metric spaces to locally compact spaces.

COROLLARY 2. 3. *Let* $\mathfrak{S}$ *be* lc^p, *L a compact subset of* $\mathfrak{S}$, *and O any neighborhood of L. Then there are at most a finite number of p-cycles in L which are independent with respect to homologies in O.*

3. Open generalized manifolds.

Definition 3. 1. A locally compact space $\mathfrak{M}$ is an open generalized manifold of dimension n if

a) $\dim \mathfrak{M} = n$,
b) $\mathfrak{M}$ is lc^n,
c) For each point $s \in \mathfrak{M}$, $R_p(s) = 0$, $p < n$,
d) For each point $s \in \mathfrak{M}$, $R_n(s) = 1$.

$\mathfrak{M}$ is said to be orientable if there is an infinite n-cycle Γ^n on $\mathfrak{M}$ which is not homologous to any cycle on any proper closed subset of $\mathfrak{M}$.

Note that any open subset of an open generalized n-manifold, or of a compact generalized n-manifold, is itself an open generalized n-manifold.

4. Cochain realizations. Let K be a finite complex of $\dim \leq n$ and let U be the nerve of a covering $\mathfrak{U}$ of the open orientable generalized manifold, $\mathfrak{M}$. A function τ^* which assigns to each chain X^q of K a cochain $\tau^*(X^q) = X_{n-q}$ of U is called a cochain realization of K on U if

a) τ^* is a linear mapping,
b) $\tau^*(FX^q) = F\tau^*X^q$, where F on the left is the boundary operator and on the right is the coboundary operator.
c) $KI(X^0) = KI(\Gamma^n \cdot \tau^* X^0)$ for every 0-chain X^0 on K.

The last condition has meaning since by b) the image under τ^* of a 0-chain (necessarily a 0-cycle) of K is an n-cocycle of U and hence of $\mathfrak{M}$.

The definition of a partial cochain realization $\tau^{*\prime}$ is now parallel to definition 1.4 of [b]. The norm of τ^* or $\tau^{*\prime}$ is defined in the obvious manner.

THEOREM 4.1. *Let $\mathfrak{M}$ be an orientable open generalized n-manifold. Let L be a compact subset of $\mathfrak{M}$ and O a neighborhood of L with a compact closure. For each covering $\mathfrak{E}$ of $\mathfrak{M}$ there is a refinement $\underline{\mathfrak{H}}_n(\mathfrak{E}; L, O)$. and for each covering $\mathfrak{B}$ of $\mathfrak{M}$ there is a refinement $\underline{\mathfrak{A}}_n(\mathfrak{B}, \mathfrak{E}; L, O)$ such that if $\tau^{*\prime}$ is a partial cochain realization on $B \cap L$ of norm $< \underline{\mathfrak{H}}_n(\mathfrak{E}; L, O)$ of an n-dimensional complex K, then there is a cochain realization $\tau^* X^q = \pi^*{}_a{}^b \tau^{*\prime} X^q$ whenever the latter is defined.*

The proof of this theorem is entirely parallel to that of Theorem 2.1 using conditions c) and d) of definition 3.1 in place of local connectedness.

5. The Poincaré duality theorem.

For open generalized manifolds this theorem takes the following form:

THEOREM 5.1. *If $\mathfrak{M}$ is an orientable open generalized n-manifold, then $H^p(\mathfrak{M}) \simeq h^{n-p}(\mathfrak{M})$.*

Proof. By Theorem L_2, $H^p(\mathfrak{M})$ is isomorphic to $h_p(\mathfrak{M})$. Let Υ_p be any compact cocycle of $\mathfrak{M}$. Then $(\Gamma^n \cdot \Upsilon_p) = \Upsilon^{n-p}$ is a compact cycle of $\mathfrak{M}$ which we denote by $\phi(\Upsilon_p)$. Clearly ϕ induces a homomorphism of $h_p(\mathfrak{M})$ into $h^{n-p}(\mathfrak{M})$. We show first that $\phi(h_p(\mathfrak{M}))$ covers $h^{n-p}(\mathfrak{M})$.

Let Υ^{n-p} be a compact cycle of $\mathfrak{M}$ and L a compact subset of $\mathfrak{M}$ which carries it. Let O_1, O_2 and O_3 be neighborhoods of L in $\mathfrak{M}$ with compact closures such that $O_1 \supset \bar{O}_2 \supset O_2 \supset \bar{O}_3$. Let $\mathfrak{E}$ be a covering of $\mathfrak{M}$ such that $St(\bar{O}_3, \mathfrak{E}) < O_2$ and choose $\mathfrak{E}_1 {}^*\!< \mathfrak{H}^n(\mathfrak{E}, \bar{O}_3, O_2)$ such that $St(L, \mathfrak{E}_1) \subset O_3$. Let $\mathfrak{U}_0$ be the covering of Theorem 2.2 corresponding to the sets O_1 and $\bar{O}_2$ and let $\mathfrak{U}_1$ be an n-dimensional refinement of $\mathfrak{U}_0$ which is ${}^* < \underline{\mathfrak{H}}_n(\mathfrak{E}_1; L, O_3)$ and $< \mathfrak{A}^n(\mathfrak{U}_0, \mathfrak{E}; \bar{O}_3, O_2)$. In each open set of $\mathfrak{U}_1$ there is a compact n-cocycle of $\mathfrak{M}$ whose intersection with Γ^n has Kronecker index equal to 1. We may assume that these cocycles each have a coordinate on a U_2 where $\mathfrak{U}_2 < \mathfrak{U}_1$. Now we define $\tau^{*\prime}$ by letting $\tau^{*\prime}$ assign to each vertex of U_1 the corresponding n-cocycle of U_2. $\tau^{*\prime}$ satisfies the conditions of Theorem 4.1, so there is a cochain realization τ^* of $U_1 \cap L$ of norm $< \mathfrak{E}_1$ on $U_3 \cap O_3$ where $\mathfrak{U}_3 = \underline{\mathfrak{A}}_n(\mathfrak{U}_2, \mathfrak{E}_1; L, O_3)$. In particular, if $x_1{}^{n-p}$ is the coordinate on U_1 of Υ^{n-p}, then $\tau^* x_1{}^{n-p} = x_p{}^3$ is a cocycle of $U_3 \cap O_3$, and $x_p{}^3$ is in O_3. Hence $x_p{}^3 = \Upsilon_p$ is a compact cocycle of $\mathfrak{M}$.

We wish to show that $(\Gamma^n \cdot \Upsilon_p) \sim \Upsilon^{n-p}$ in O_1. To this end, let K be the product of $U_1 \cap \bar{O}_3$ with the unit interval, simplicially subdivided so that all the vertices of K are on the base and the top of K. We identify $U_1 \cap \bar{O}_3$ with the base of K. Note that if x^q is a cycle of $U_1 \cap \bar{O}_1$ and if $\bar{x}^q$ is the corresponding cycle on the top of K, then there is a chain y^{q+1} in K such that $Fy^{q+1} = x^q - \bar{x}^q$.

We define now a partial realization τ' of K as follows: on the base of K, τ' is the identity mapping and for any chain $\bar{x}^q$ on the top of K, $\tau'\bar{x}^q = (-1)^{nq}\pi_3{}^1(X_3{}^n \cdot \tau^* x^q)$, where $X_3{}^n$ is a coordinate of Γ^n on U_3. To show that τ' is a partial realization it is only necessary to show that $\tau' F\bar{x}^q = F\tau'\bar{x}^q$ (see definition 1.2 of [B]).[7] But we have, for any $\bar{x}^q$ on the top of K,

$$\begin{aligned} F\tau'\bar{x}^q = F[(-1)^{nq}\pi_3{}^1(X_3{}^n \cdot \tau^* x^q)] &= (-1)^{nq}\pi_3{}^1 F(X_3{}^n \cdot \tau^* x^q) \\ &= (-1)^{nq}\pi_3{}^1[(FX_3{}^n \cdot \tau^* x^q) + (-1)^n (X_3{}^n \cdot F\tau^* x^q)] \\ &= (-1)^{n(q+1)}\pi_3{}^1(X^n \cdot \tau^* F x^q); \end{aligned}$$

Also, $\tau' F x^q = (-1)^{n(q-1)}\pi_3{}^1(X_3{}^n \cdot \tau^* F x^q)$. Since $n(q+1)$ and $n(q-1)$ are odd or even together, τ' is a chain mapping and hence a partial realization. Since $\mathfrak{E}_1 {}^*{<'}\, \mathfrak{H}^n(\mathfrak{E}, \bar{O}_3, O_2)$, norm $\tau' < \bar{\mathfrak{H}}^n(\mathfrak{E}; \bar{O}_3, O_2)$. Therefore, by the choice of $\mathfrak{U}_1 < \bar{\mathfrak{A}}^n(\mathfrak{U}_0, \mathfrak{E}; \bar{O}_3, O_2)$, τ' can be extended to a realization τ of K on $(U_1 \cup U_0) \cap O_2$.

Now let y^{n-p-1} be a chain on K such that $Fy^{n-p-1} = x_1{}^{n-p} - \bar{x}_1{}^{n-p}$, where $x_1{}^{n-p}$ is the coordinate of Υ^{n-p} on U_1. Then $F\tau y^{n-p-1} = x_1{}^{n-p} - \pi_3{}^1(X_3{}^n \cdot \tau^* x_1{}^{n-p})$ and τy^{n-p-1} is on $(U_1 \cup U_0) \cap O_2$. Let π be the natural mapping of $(U_1 \cup U_0)$ into U_0. Then $F\pi\tau y^{n-p-1} = \pi_1{}^0 x_1{}^{n-p} - \pi_1{}^0\pi_3{}^1(X_3{}^n \cdot \tau^* x_1{}^{n-p})$. Thus the coordinates of Υ^{n-p} and $(\Gamma^n \cdot \Upsilon_p)$ on U_0 are homologous on $\bar{O}_2$. But by Theorem 2.2, this is sufficient to show that $(\Gamma^n \cdot \Upsilon_p) \sim \Upsilon^{n-p}$ in $\bar{O}_1$. Hence ϕ is a homomorphism of $H^p(\mathfrak{M})$ onto $h^{n-p}(\mathfrak{M})$.

To complete the proof, it is sufficient to show that there is a homomorphism ψ of $h^{n-p}(\mathfrak{M})$ onto $H^p(\mathfrak{M})$. The existence of such a homomorphism is demonstrated by the same methods as used above but starting with $h^{n-p}(\mathfrak{M})$ and Theorem L_1. The details are left to the reader.

Remark. The above proof shows that the two groups involved have the same number of generators. Since the coefficient group is a field, this means that the two groups are isomorphic. However, since each one of the groups has a finite basis, it is easy to see that the homomorphism ϕ constructed above is itself an isomorphism.

---

[7] In a similar construction in the proof of Theorem 7.3 of [B], the factor $(-1)^{nq}$ was omitted. It should be inserted there.

In the above proof we have assigned to each compact cycle Υ^{n-p} a compact cocycle Υ_p, which we denote by $\tau^*\Upsilon^{n-p}$, in such a way that $(\Gamma^n \cdot \tau^*\Upsilon^{n-p}) \sim \Upsilon^{n-p}$. If $\Upsilon^{n-p} \sim 0$ on a set L_1, then there is a chain y^{n-p-1} on $U_1 \cap L_1$ (same notation as above) such that $Fy^{n-p-1} = x_1^{n-p}$. Then $\tau^* y^{n-p-1} = y^3{}_{p+1}$ is a cochain on U_3 such that $Fy^3{}_{p+1} = \tau^* F y_1{}^{n-p-1} = \tau^* x_1{}^{n-p} = x_p{}^3$. Hence $\tau^*\Upsilon^{n-p} \sim 0$, and it is clear that this cohomology can be made to take place in any assigned neighborhood of L_1. We gather these results in the following lemma for future use.

LEMMA 5. 2. *Let* $\mathfrak{M}$ *be an orientable open generalized n-manifold and let* Υ^{n-p} *be a compact cycle on a compact set* L *and* ~ 0 *on a compact set* L_1. *Then for any choice of neighborhoods* O *and* O_1 *of* L, L_1, *there is a compact cocycle* $\Upsilon_p = \tau^*\Upsilon^{n-p}$ *in* O *such that* $(\Gamma^n \cdot \Upsilon_p) \sim \Upsilon^{n-p}$ *in* O *and* $\Upsilon_p \sim 0$ *in* O_1.

6. The Alexander and other duality theorems. In Section **6** of [B] we reproduced Čech's definition of local Betti numbers and also gave a definition of local co-Betti numbers. Now suppose that S is a closed subset of a space $\mathfrak{X}$ and that s is a point of S. Then there are local Betti and co-Betti numbers of the complement of S at the point s. These are defined in the same way as the local Betti and co-Betti numbers of S at the point s but with the neighborhoods P and Q of s replaced by $P - S$ and $Q - S$. These numbers are denoted by $R^q(\mathfrak{X} - S, s)$ and $R_q(\mathfrak{X} - S, s)$. We shall hereafter use $R^q(S, s)$ and $R_q(S, s)$ to denote the local Betti and co-Betti numbers of S at s.

THEOREM 6. 1. *For each point* s *of* S, $R^q(\mathfrak{X} - S, s) = R_q(\mathfrak{X} - S, s)$.

The proof is exactly the same as that of Theorem 6. 2 of [B].

THEOREM 6. 2. *If* $R_q(\mathfrak{X}, s) = R_{q+1}(\mathfrak{X}, s) = 0$, *then* $R_{q+1}(\mathfrak{X} - S, s) = R_q(S, s)$.

Proof. We first show that $R_q(S, s) \leq R_{q+1}(\mathfrak{X} - S, s)$. Let O be any neighborhood of s in $\mathfrak{X}$. Since $R_q(\mathfrak{X}, s) = 0$, O contains a neighborhood P of s such that $R_q(O, P, s) = 0$. Let Q be any neighborhood of s which is in P.

Now let $\Gamma_q{}^1, \cdots, \Gamma_q{}^k$ be cocycles of S in $Q \cap S$ which are independent in $O \cap S$. Let $\mathfrak{U}_\zeta$ be a covering of $\mathfrak{X}$ such that each $\Gamma_q{}^i$ has a coordinate $X_q{}^{i\zeta}$ on U_ζ. Then $FX_q{}^{i\zeta}$ is a $(q+1)$-dimensional cocycle in $Q - S$. Let us suppose that these cocycles are not independent in $P - S$. Then there is a chain $Y_q{}^\rho$ on U_ρ, for some $\mathfrak{U}_\rho < \mathfrak{U}_\zeta$, such that $Y_q{}^\rho$ is in $P - S$ and $FY_q{}^\rho = \pi^*{}_\rho{}^\zeta \Sigma f_i (FX_q{}^{i\zeta})$. Then $(Y_q{}^\rho - \pi^*{}_\rho{}^\zeta \Sigma f_i X_q{}^{i\zeta})$ is a cocycle of $\mathfrak{X}$ in P.

By the choice of P, this cocycle must bound in O. Hence there is a chain $Z^{\mu}{}_{q-1}$ on U_μ, for some $\mathfrak{U}_\mu < \mathfrak{U}_\rho$, such that $Z^{\mu}{}_{q-1}$ is in O and $FZ^{\mu}{}_{q-1} = \pi^*{}_{\mu}{}^{\rho} Y_q{}^{\rho} - \pi^*{}_{\mu}{}^{\rho}\pi^*{}_{\rho}{}^{\zeta}\Sigma f_i X_q{}^{i\zeta}$. But since $Y_q{}^{\rho}$ is in $O - S$, we have $\Sigma f_i \Gamma_q{}^i \sim 0$ in $O \cap S$, which is impossible. Therefore the cocycles $FX_q{}^{i\zeta}$ are independent in $P - S$, and hence we have $R_q(O \cap S, Q \cap S, s) \leq R_{q+1}(P - S, Q - S, s)$. From this we have immediately that $R_q(S, s) \leq R_{q+1}(\mathfrak{X} - S, s)$.

To prove the reverse inequality, let O be an arbitrary neighborhood of s. Since $R_{q+1}(\mathfrak{X}, S) = 0$, there are neighborhoods $P \supset Q$ of s contained in O such that $R_{q+1}(P, Q, s) = 0$.

Now let $\Gamma^1{}_{q+1}, \cdots, \Gamma^k{}_{q+1}$ be cocycles in $Q - S$ which are independent in $O - S$. By the choice of Q with relation to P, each of these cocycles bounds in P. Hence there is a covering $\mathfrak{U}_\zeta$ of $\mathfrak{X}$ such that each $\Gamma^i{}_{q+1}$ has a coordinate $X^{i\zeta}{}_{q+1}$ on U_ζ, and there are chains $Y_q{}^{i\zeta}$ on U_ζ such that $Y_q{}^{i\zeta}$ is in P and $FY_q{}^{i\zeta} = X^{i\zeta}{}_{q+1}$. Let $Y_q{}^{i\zeta} = A_q{}^{i\zeta} + B_q{}^{i\zeta}$, where $A_q{}^{i\zeta}$ is the part of $Y_q{}^{i\zeta}$ in $\mathfrak{X} - S$. Then each $B_q{}^{i\zeta}$ is a q-dimensional cocycle of S in $P \cap S$. Let us suppose that these cocycles are not independent in $O \cap S$. Then on some U_ρ, $\mathfrak{U}_\rho < \mathfrak{U}_\zeta$, there is a chain $Z^{\rho}{}_{q-1}$ in O such that $FZ^{\rho}{}_{q-1} = \pi^*{}_{\rho}{}^{\zeta}\Sigma f_i B_q{}^{i\zeta} + C_q{}^{\rho}$, where $C_q{}^{\rho}$ is in $O - S$. Then we have $0 = FFZ^{\rho}{}_{q-1} = FC_q{}^{\rho} + \pi^*{}_{\rho}{}^{\zeta}\Sigma f_i(FB_q{}^{i\zeta}) = FC_q{}^{\rho} + \pi^*{}_{\rho}{}^{\zeta}\Sigma f_i(F(Y_q{}^{i\zeta} - A_q{}^{i\zeta}))$. Then $\pi^*{}_{\rho}{}^{\zeta}\Sigma f_i FY_q{}^{i\zeta} = \pi^*{}_{\rho}{}^{\zeta}\Sigma f_i X^{i\zeta}{}_{q+1} = F(C_q{}^{\rho} + \pi^*{}_{\rho}{}^{\zeta}\Sigma f_i A_q{}^{i\zeta})$. But $C_q{}^{\rho}$ is in $O - S$ and so is each $A_q{}^{i\zeta}$. Hence $\Sigma f_i \Gamma_{q+1} \sim 0$ in $O - S$, which is impossible. Therefore the cocycles $B_q{}^{i\zeta}$ are independent in $O \cap S$, so we have $R_{q+1}(O - S, Q - S, s) \leq R_q(O \cap S, P \cap S, s)$. From this we have immediately that $R_{q+1}(\mathfrak{X} - S, s) \leq R_q(S, s)$ which completes the proof.

THEOREM 6.3.[8] *If* $h_q{}^a(\mathfrak{X}) = h^a{}_{q+1}(\mathfrak{X}) = 0$, *then* $h_q{}^a(S) = h^a{}_{q+1}(\mathfrak{X} - S)$.

Proof. Let $\Upsilon_q{}^1, \cdots, \Upsilon_q{}^k$ be independent compact cocycles of S, and let $\Gamma_q{}^i$ be a Čech cocycle which is a component of $\Upsilon_q{}^i$. Let G be an open set whose closure in $\mathfrak{X}$ is compact and which carries all the $\Gamma_q{}^i$. Repeating the construction in the first part of the proof of the above theorem, we obtain the cocycles $FX_q{}^{i\zeta}$. By a suitable choice of $\mathfrak{U}_\rho$, these cocycles all lie in an open set $G' \subset \mathfrak{X} - S$ whose closure in $\mathfrak{X}$ is compact. Hence we may use Theorem 1.1 to obtain compact cocycles $x^i{}_{q+1} \sim FX_q{}^{i\zeta}$ in G'. Also, by the argument above taken in conjunction with Theorem 1.1, these compact cocycles are independent in $\mathfrak{X} - S$. Thus rank $(h_q{}^a(S)) \leq$ rank $(h^a{}_{q+1}(\mathfrak{X} - S))$. The reverse inequality is proved in an analogous fashion.

---

[8] See pp. 227-229 of [15]. Also, see P. Alexandroff, "General combinatorial topology," *Transactions of the American Mathematical Society*, vol. 49 (1941), pp. 41-105, where this theorem is called Kolmogoroff's duality law.

As in the case of Theorem 5.1, the particular construction used above leads to an isomorphism since each one of the two groups has a finite basis.

We are now ready to prove our main theorem.

THEOREM 6.4. [Alexander's Duality Theorem]. *Let $\mathfrak{M}$ be an orientable open connected n-dimensional generalized manifold which has the homology groups of the n-sphere. If L is any closed subset of $\mathfrak{M}$, all of whose Betti numbers are finite, then* $H_a^q(L) \cong h_a^{n-q-1}(\mathfrak{M} - L)$ *for* $q \leq n - 1$.

Proof. Since $\mathfrak{M}$ is connected and n-sphere-like, $H_a^q(\mathfrak{M}) = 0$ for $q \leq n - 1$. Hence $h_q^a(\mathfrak{M}) = 0$ for $q \leq n - 1$. Therefore, by Theorem 6.3, $H_a^q(L) \simeq h_q^a(L) \simeq h^a_{q+1}(\mathfrak{M} - L)$ for $q < n - 1$. Since $q + 1 > 0$, $h^a_{q+1}(\mathfrak{M} - L) = h_{q+1}(\mathfrak{M} - L)$, and by Theorem 5.1, $h_{q+1}(\mathfrak{M} - L) \simeq h^{n-q-1}(\mathfrak{M} - L)$. Since $q < n - 1$, $n - q - 1 > 0$, so $h^{n-q-1}(\mathfrak{M} - L) \simeq h_a^{n-q-1}(\mathfrak{M} - L)$, which completes the proof for $q < n - 1$.

For $q = n - 1$, let $\Upsilon^1_{n-1}, \cdots, \Upsilon^k_{n-1}$ be compact cocycles of L which are independent on L. As in the first part of the proof of Theorem 6.3, there are compact cocycles $\Upsilon_n^1, \cdots, \Upsilon_n^k$ in $\mathfrak{M} - L$ which are independent in $\mathfrak{M} - L$. By Theorem 5.1, to each Υ_n^i there corresponds a compact cycle Υ_i^0 of $\mathfrak{M} - L$ in such a way that the Υ_i^0 are independent in $\mathfrak{M} - L$ and furthermore $KI(\Upsilon_i^0) = KI(\Gamma^n \cdot \Upsilon_n^i)$. But since each Υ_n^i is ~ 0 in $\mathfrak{M}$, $KI(\Gamma^n \cdot \Upsilon_n^i) = 0$. Hence the Υ_i^0 determine independent elements of $h_a^0(\mathfrak{M} - L)$ and so rank $(H_a^{n-1}(L)) \leq$ rank $(h_a^0(\mathfrak{M} - L))$.

Conversely, let $\Upsilon_1^0, \cdots, \Upsilon_k^0$ be independent compact cycles of $\mathfrak{M} - L$ with $KI(\Upsilon_i^0) = 0$ for each i. By Theorem 5.1, there are compact cocycles Υ_n^i of $\mathfrak{M} - L$, such that $(\Gamma^n \cdot \Upsilon_n^i) \sim \Upsilon_i^0$. Hence $KI(\Gamma^n \cdot \Upsilon_m^i) = 0$. It follows from Lefschetz's Theorem ([B] p. 568) that each $\Upsilon_m^i \sim 0$ in $\mathfrak{M}$. Now the second part of the proof of Theorem 6.2 together with Theorem 1.1 applies to show that there are at least k compact $(n - 1)$-cocycles of L which are independent on L. Hence rank $(h_a^0(\mathfrak{M} - L)) \leq$ rank $(H_a^{n-1}(L))$, and this is sufficient to show that $H_a^{n-1}(L) \simeq h_a^0(\mathfrak{M} - L)$, which completes the proof.

Čech [7] has shown that the above theorem can be localized in the following manner: Let $\mathfrak{T}$ be a locally compact space and S a closed subset of $\mathfrak{T}$. Let $P \supset Q$ be neighborhoods of a point s of S. Denote by $r^q(P - S, Q - S)$ the maximum number of compact q-cycles in $Q - S$ which are independent with respect to homologies in $P - S$. Then, using these numbers, define $r^q(\mathfrak{T} - S, s)$ in the obvious way.

THEOREM 6.5. *Let $\mathfrak{M}$ be an open generalized n-manifold and let L be a closed subset of $\mathfrak{M}$. Then for each point s of L,* $R^q(L, s) = r^{n-q-1}(\mathfrak{M} - L, s)$ *for all* $q \leq n - 1$.

Proof. Let $P \supset Q$ be neighborhoods of s such that $R_q(P, Q) = R_{q+1}(P, Q) = 0$ and $R_n(P, Q) = 1$. Then, since L is closed in $\mathfrak{M}$, $P - L$ and $Q - L$ are open orientable generalized n-manifolds. Now, by trivial modifications in the proofs of Theorems 6.2 and 5.1, we have $R^q(P \cap L, Q \cap L) = R_q(P \cap L, Q \cap L) = R_{q+1}(P - L, Q - L) = r^{n-q-1}(P - L, Q - L)$. This proves the theorem.

As an immediate consequence we have the following result on uniform local connectedness.[9]

COROLLARY 6.6. *Let $\mathfrak{M}$ be a compact generalized n-manifold and let L be a closed subset of $\mathfrak{M}$. If $R^q(L, s) = 0$ for every point s of L, then $\mathfrak{M} - L$ is $(n - q - 1)$ — ulc.*

7. **Geometric linking.** Let S be a closed subset of a compact space $\mathfrak{T}$. A cycle Γ^p of S and a compact cycle Υ^q of $\mathfrak{T} - S$ are said to be *linked* if neither one is homologous to zero (as a compact cycle) in the complement of any carrier[5] of the other. If (Γ^p) is the element of $H^p(S)$ containing Γ^p and (Υ^q) the element of $h^q(\mathfrak{T} - S)$ containing Υ^q, we say that (Γ^p) and (Υ^q) are linked if any element of one class is linked with any element of the other.

THEOREM 7.1. *Let $\mathfrak{M}$ be a compact orientable generalized n-manifold which is n-sphere-like, and let L be a closed subset of $\mathfrak{M}$ all of whose Betti numbers are finite. Then in the isomorphism of $H^p(L)$ and $h^{n-p-1}(\mathfrak{M} - L)$ established in Theorem 6.4, corresponding elements are linked.*

Proof.[10] If we trace the isomorphisms involved in the proof of Theorem 6.4, we see that to each element (Γ^p) of $H^p(L)$ there is assigned an element (Υ^{n-p-1}) of $h^{n-p-1}(\mathfrak{M} - L)$ in the following way: an arbitrary cycle Γ^p is chosen in (Γ^p) and a cocycle Γ_p of L is given by Theorem L_1 such that $KI(\Gamma^p \cdot \Gamma_p) \neq 0$. Then $\{F\Gamma_p{}^\xi\} = \{A^\xi{}_{p+1}\}$ is a cocycle of $\mathfrak{M} - L$. By Theorem 1.1, $\{A^\xi{}_{p+1}\}$ is cohomologous to a compact cocycle Υ_{p+1} of $\mathfrak{M} - L$. Then $(\Gamma^n \cdot \Upsilon_{p+1}) = \Upsilon^{n-p-1}$ is a compact cycle of $\mathfrak{M} - L$ and the class (Υ^{n-p-1}) is the one assigned to (Γ^p).

Now let Γ^p be any cycle of L and let Υ^{n-p-1} be the compact cycle of $\mathfrak{M} - L$ which is obtained from Γ^p by the above process, and let $\bar{\Upsilon}^{n-p-1}$ be any element of (Υ^{n-p-1}). If we can show that Γ^p and Υ^{n-p-1} are linked, the theorem will be proved.

---

[9] Cf. Theorem 1 of [20].

[10] Cf. L. Pontrjagin, "Zum Alexanderschen Dualitätssatz," *Göttingen Nachrichten* (1927), pp. 315-329. I am indebted to Professor R. L. Wilder both for the suggestion that this theorem might hold and for an outline of essentially the proof used here.

First let us suppose that there is a carrier $[\bar{\Upsilon}^{n-p-1}]$ of $\bar{\Upsilon}^{n-p-1}$ such that $\Gamma^p \sim 0$ on a compact set T which does not meet $[\bar{\Upsilon}^{n-p-1}]$. Then there is a neighborhood O of $[\bar{\Upsilon}^{n-p-1}]$ whose closure is compact, is contained in $\mathfrak{M} - L$, and does not meet T. By Lemma 5.2, there is a compact cocycle $\bar{\Upsilon}_{p+1} = \tau^* \bar{\Upsilon}^{n-p-1}$ in O such that $(\Gamma^n \cdot \bar{\Upsilon}_{p+1}) \sim \bar{\Upsilon}^{n-p-1}$ in O. Since $H^p(L)$ has a finite basis, the homomorphism $\Upsilon_{p+1} \rightarrow (\Gamma^n \cdot \Upsilon_{p+1})$ is an isomorphism. Therefore, since $\Upsilon^{n-p-1} \sim \bar{\Upsilon}^{n-p-1}$ in $\mathfrak{M} - L$, $\Upsilon_{p+1} \sim \bar{\Upsilon}_{p+1}$ in $\mathfrak{M} - L$, where Υ_{p+1} is the compact cocycle above such that $\Upsilon^{n-p-1} = (\Gamma^n \cdot \Upsilon_{p+1})$.

Let $\mathfrak{U}_\zeta$ be a covering of M such that Γ_p has a coordinate on U_ζ. Let $\mathfrak{U}_\rho$ be a refinement of $\mathfrak{U}_\zeta$ such that $\pi^*{}_\rho{}^\zeta F \Gamma_p{}^\zeta = \pi^*{}_\rho{}^\zeta A^\zeta{}_{p+1} \sim \Upsilon^\rho{}_{p+1}$ on U_ζ and such that $\Upsilon^\rho{}_{p+1} \sim \bar{\Upsilon}^\rho{}_{p+1}$ on U_ρ. Let $B_p{}^\rho$ in $\mathfrak{M} - L$ be such that $FB_p{}^\rho = \pi^*{}_\rho{}^\zeta A^\zeta{}_{p+1} - \Upsilon^\rho{}_{p+1}$, and let $C_p{}^\rho$ in $\mathfrak{M} - L$ be such that $FC_p{}^\rho = \bar{\Upsilon}^\rho{}_{p+1} - \Upsilon^\rho{}_{p+1}$. Then $F(\pi^*{}_\rho{}^\zeta \Gamma_p{}^\zeta - B_p{}^\rho + C_p{}^\rho) = \bar{\Upsilon}^\rho{}_{p+1}$. Since $[\bar{\Upsilon}_{p+1}]$ does not meet T, the part of $(\pi^*{}_\rho{}^\zeta \Gamma_p{}^\zeta - B_p{}^\rho + C_p{}^\rho)$ on $L \cup T$ is a cocycle $\bar{\Gamma}_p$ of $L \cup T$. Also, $KI(\Gamma^p \cdot \bar{\Gamma}_p) = KI(\Gamma_\rho{}^p \cdot \Gamma_p{}^\rho) = KI(\Gamma_\rho{}^p \cdot \pi^*{}_\rho{}^\zeta \Gamma_p{}^\zeta) = KI(\Gamma^p \cdot \Gamma_p)$, since $B_p{}^\rho$ and $C_p{}^\rho$ do not meet L. But $KI(\Gamma^p \cdot \Gamma_p) \neq 0$ while $KI(\Gamma^p \cdot \Gamma_p) = 0$, since $\Gamma^p \sim 0$ on $L \cup T$ and consequently has Kronecker index zero with every cocycle of $L \cup T$. This contradiction shows that Γ^p cannot bound in the complement of any carrier of $\bar{\Upsilon}^{n-p-1}$.

On the other hand, suppose that there is a closed set T such that $\bar{\Upsilon}^{n-p-1}$ bounds on T while T does not meet some carrier $[\Gamma^p]$ of Γ^p. Let O be an open set containing T with the closure of O not meeting $[\Gamma^p]$. By Lemma 5.2, there is a cocycle $\bar{\Upsilon}_{p+1}$ in $O \cap (\mathfrak{M} - L)$ such that $\bar{\Upsilon}_{p+1} \sim \Upsilon_{p+1}$ in $\mathfrak{M} - L$ as above. Also, since $\bar{\Upsilon}^{n-p-1} \sim 0$ in T, $\bar{\Upsilon}_{p+1} \sim 0$ in O. Now let $\mathfrak{U}_\zeta$, $\mathfrak{U}_\rho$, $B_p{}^\rho$, and $C_p{}^\rho$ be as above. We may also assume that on U_ρ there is a chain $D_p{}^\rho$ in O such that $FD_p{}^\rho = \bar{\Upsilon}^\rho{}_{p+1}$. Let the part of $D_p{}^\rho$ on L be $\bar{D}_p{}^\rho$, so that $\bar{D}_p{}^\rho$ is a coordinate of a cocycle $\bar{\Gamma}_p$ of L. Now $(\pi^*{}_\rho{}^\zeta \Gamma_p{}^\zeta - B_p{}^\rho + C_p{}^\rho) - D_p{}^\rho$ is a cocycle of U_ρ and hence of $\mathfrak{M}$. Since $\mathfrak{M}$ is n-sphere-like, this cocycle must be cohomologous to zero in $\mathfrak{M}$. Therefore there is a $\mathfrak{U}_\mu < \mathfrak{U}_\rho$ such that $\pi^*{}_\mu{}^\rho (\pi^*{}_\rho{}^\zeta \Gamma_p{}^\zeta - B_p{}^\rho + C_p{}^\rho) - \pi^*{}_\mu{}^\rho D_p{}^\rho = FE^\mu{}_{p-1}$. The part of $E^\mu{}_{p-1}$ on L is a chain $\bar{E}^\mu{}_{p-1}$ such that the part of $F\bar{E}^\mu{}_{p-1}$ which is on L is just $\pi^*{}_\mu{}^\rho \pi^*{}_\rho{}^\zeta \Gamma_p{}^\zeta - \pi^*{}_\mu{}^\rho \bar{D}_p{}^\rho$. Hence $\Gamma_p \sim \bar{\Gamma}_p$ on L, and consequently $KI(\Gamma^p \cdot \Gamma_p) = KI(\Gamma^p \cdot \bar{\Gamma}_p) \neq 0$. But $\bar{\Gamma}_p$ is in O, which does not meet $[\Gamma^p]$, so that the intersection of Γ^p and $\bar{\Gamma}_p$ must be vacuous. This contradiction completes the proof of the theorem.

YALE UNIVERSITY.

THE CONTACT OF A CUBIC SURFACE WITH A RULED SURFACE.*

By J. Ernest Wilkins, Jr.

1. Introduction. If a surface S in three-dimensional projective space is nonruled, it is known (see Lane [1]) that there exists a four-parameter family of cubic surfaces having contact of order four at a general point of S. The only nonruled surfaces, at each point of which there exists a single cubic surface with contact of order five, are the cubic surfaces, and at each point of a cubic surface there is a one-parameter family of cubic surfaces having contact of order five. It is the purpose of this paper to investigate these questions for ruled surfaces. We find it convenient to separate the discussion into two parts, according as S is developable or not. In the first case we shall prove the following

Theorem I. *At a general point of a developable surface, not a plane or a cone, there is a three-parameter family of cubic surfaces with contact of order five. There is a unique cubic surface with contact of order six and no cubic surface with contact of order seven.*

In case S is nondevelopable we shall prove the following

Theorem II. *At a general point of a nondevelopable ruled surface S, there is a one-parameter family of cubic surfaces with contact of order five. The only such surfaces for which at each point there exists a cubic surface with contact of order six are the cubic ruled surfaces and for these surfaces S the only cubic surface with sixth-order contact is the surface S itself.*

In order to prove these theorems we shall develop power series expansions for the surface S. These expansions are obtained from a system of partial differential equations which S must satisfy. It is hoped that other applications of these series will be made on another occasion.

I. Developable Surfaces.

2. The differential equations. It is known (see Lane [2, p. 98]) that

---

* Received October 21, 1943; Presented to the American Mathematical Society, April 28, 1944.

the parametric vector equation of a developable surface not a plane or a cone can be taken in the form

$$x(u,v) = y'(v) + uy(v), \tag{2.1}$$

where $y(v)$ is the edge of regression of the developable and is a skew curve. Moreover, we may assume that the proportionality factor and the variable v have been chosen so that $y(v)$ satisfies the Laguerre-Forsythe canonical differential equation (see Lane [2, p. 84])

$$y''' = cy' - dy, \tag{2.2}$$

the quantities c and d being functions of v. Then it is easy to see that $x(u,v)$ is a solution of the completely integrable system of partial differential equations

$$\begin{aligned} &x_{uu} = 0, \quad x_{uv} = x - ux_u, \\ &x_{vvv} = (u^3 + c)x - (u^4 + cu + d)x_u - u^2x_v + ux_{vv}. \end{aligned} \tag{2.3}$$

Conversely, if $x(u,v)$ is a solution of these three equations, the first two imply that x has the form (2.1) and then the third implies that y satisfies (2.2).

3. The power series expansions. Since S is not a plane, the points x, x_u, x_v, x_{vv} may be chosen as the vertices of a local tetrahedron of reference at the point x with the unit point so selected that a point $X = x_1x + x_2x_u + x_3x_v + x_4x_{vv}$ will have local homogeneous coördinates proportional to (x_1, x_2, x_3, x_4). If $X = x(u + \Delta u, v + \Delta v)$ is a point on S near x, we can employ a Taylor series and the differential equations (2.3) in a familiar fashion to express each of the local coördinates of X as a power series in Δu and Δv. These series, to as many terms as will be needed in the subsequent discussion, are as follows:

$$\begin{aligned} x_1 &= 1 + \Delta u\Delta v - \tfrac{1}{2}u\Delta u\Delta v^2 + \tfrac{1}{6}(u^3 + c)\Delta v^3 + \tfrac{1}{6}u^2\Delta u\Delta v^3 \\ &\quad + \tfrac{1}{24}(c' - d)\Delta v^4 + \cdots, \\ x_2 &= \Delta u - u\Delta u\Delta v + \tfrac{1}{2}u^2\Delta u\Delta v^2 - \tfrac{1}{6}(u^4 + cu + d)\Delta v^3 + \cdots, \\ x_3 &= \Delta v + \tfrac{1}{2}\Delta u\Delta v^2 - \tfrac{1}{6}u^2\Delta v^3 - \tfrac{1}{6}u\Delta u\Delta v^3 + \tfrac{1}{24}c\Delta v^4 \\ &\quad + \tfrac{1}{120}(2c' - d)\Delta v^5 + \cdots, \\ x_4 &= \tfrac{1}{2}\Delta v^2 + \tfrac{1}{6}u\Delta v^3 + \tfrac{1}{6}\Delta u\Delta v^3 + \tfrac{1}{120}c\Delta v^5 \\ &\quad + \tfrac{1}{720}(3c' - d + cu)\Delta v^6 + \cdots. \end{aligned} \tag{3.1}$$

If local nonhomogeneous coördinates are introduced by placing

$$x = x_2/x_1, \qquad y = x_3/x_1, \qquad z = x_4/x_1,$$

we find that they may be expressed in power series as follows:

$$\begin{aligned} x &= \Delta u - u\Delta u\Delta v - \Delta u^2\Delta v + \tfrac{1}{2}u^2\Delta u\Delta v^2 - \tfrac{1}{6}(u^4 + cu + d)\Delta v^3 + \cdots, \\ y &= \Delta v - \tfrac{1}{2}\Delta u\Delta v^2 - \tfrac{1}{6}u^2\Delta v^3 + \tfrac{1}{3}u\Delta u\Delta v^3 - \tfrac{1}{24}(4u^3 + 3c)\Delta v^4 \\ &\quad + \tfrac{1}{2}\Delta u^2\Delta v^3 + \tfrac{1}{120}(4d - 3c')\Delta v^5 + \cdots, \\ z &= \tfrac{1}{2}\Delta v^2 + \tfrac{1}{6}u\Delta v^3 - \tfrac{1}{3}\Delta u\Delta v^3 + \tfrac{1}{12}u\Delta u\Delta v^4 - \tfrac{1}{120}(10u^3 + 9c)\Delta v^5 \\ &\quad + \tfrac{1}{3}\Delta u^2\Delta v^4 - \tfrac{1}{720}(12c' - 14d + 19cu + 20u^4)\Delta v^6 + \cdots. \end{aligned} \tag{3.2}$$

It is now possible to calculate an expansion for z as a power series in x and y. Thus we obtain

$$\begin{aligned} z &= \tfrac{1}{2}y^2 + \tfrac{1}{6}uy^3 + \tfrac{1}{6}xy^3 + \tfrac{1}{6}u^2y^4 + \tfrac{1}{6}uxy^4 + \tfrac{1}{60}(10u^3 + 3c)y^5 \\ &\quad + \tfrac{1}{8}x^2y^4 + \tfrac{1}{4}u^2xy^5 + \tfrac{1}{360}(3c' + 5d + 65u^4 + 23cu)y^6 + \cdots. \end{aligned} \tag{3.3}$$

We conclude that *the expansion* (3.3) *for one nonhomogeneous coördinate z as a power series in the other two nonhomogeneous coördinates x and y of a point represents an integral surface of equations* (2.3) *in a sufficiently small neighborhood of an ordinary point x whose local nonhomogeneous coördinates are* $0, 0, 0$, *the vertices of the local tetrahedron of reference being the points x, x_u, x_v, x_{vv} and the unit point being suitably chosen.*

4. Contact of a quadric surface with a developable surface. Before taking up the proof of Theorem I, it is of some interest to discuss quadric surfaces. Writing the equation of the most general quadric surface and demanding that it be satisfied by the series (3.3) identically in x and y as far as terms of the second degree, we find that there exists a three-parameter family of quadric surfaces with second-order contact. The equation of a general one of these is

$$z - \tfrac{1}{2}y^2 + k_2xz + k_3yz + k_4z^2 = 0, \tag{4.1}$$

where k_2, k_3 and k_4 are the parameters. The quadric surface has third-order contact if, and only if, $k_2 = 0$, $k_3 = -u/3$. We conclude that there exists a one-parameter family of quadric surfaces having third-order contact. The equation of a general one of these is

$$z - \tfrac{1}{2}y^2 - uyz/3 + k_4z^2 = 0, \tag{4.2}$$

where k_4 is the parameter. It is interesting to note that the quadrics (4.2) are all quadric cones with their vertex at the focal point on the generator through the origin. Finally, we find that the quadric (4.2) never has fourth-order contact for any value of k_4.

5. Contact of a cubic surface with a developable surface. We now proceed to the proof of Theorem I. We write the equation of the most general cubic surface and demand that it be satisfied by the power series (3. 3) identically in x and y as far as terms of the fifth degree. Thus we find that each member of the three-parameter family of cubic surfaces whose equations are

$$(5.1)\quad \begin{aligned} &(9k_1 + 15k_3)y^3 + k_4 z^3 + (12k_1 + 40uk_3)xz^2 + 30k_3 xyz \\ &\quad + [2uk_2 + 8u^2k_1 + (40u^3 + 18c)k_3]yz^2 - 6k_2 z^2 - 18k_1 yz \\ &\quad + (6uk_1 + 3k_2 + 30u^2k_3)y^2 z - 90k_3 z + 45k_3 y^2 = 0, \end{aligned}$$

has fifth-order contact with S. Moreover, there is a unique member of this family which has sixth-order contact with S. Its equation is

$$(5.2)\quad 135y^3 + (80u^3 + 108c)z^3 + 180xz^2 + 120u^2yz^2 + 90uy^2z - 270yz = 0.$$

This cubic surface could be called the *osculating cubic surface* of the developable surface. Finally, the cubic surface (5. 2) never has contact of the seventh order. These remarks complete the proof of Theorem I.

We have incidentally proved that there are no cubic surfaces which are also tangent developables of twisted curves.

6. The degenerate developables. Let us now consider the cases that arise when the developable surface S is a plane or a cone. In the first case we may assume that the equation of the plane is $z = 0$. Then it is easy to see that there exists a thirteen-parameter family of cubic surfaces with second-order contact and no non-composite cubic surface with contact of the third order.

If the developable surface S is a cone and P is a point on S at which we wish to investigate the order of contact of cubic surfaces, we may suppose that the director curve $y(v)$ is a plane curve through P and that the proportionality factor and variable v have been chosen so that $y(v)$ is a solution of the Laguerre-Forsythe canonical differential equation

$$(6.1)\qquad y''' = py.$$

Let $z = 0$ be the plane of this curve and pick the points y, y', $\frac{1}{2}y''$ as three of the vertices of a local tetrahedron of reference. The fourth vertex of the tetrahedron is picked as the vertex a of the cone and the unit point is selected so that a point $X = x_1 y + x_2 y' + \frac{1}{2}x_3 y'' + x_4 a$ will have local homogeneous coördinates proportional to (x_1, x_2, x_3, x_4). Then the power series expansion for the cone analogous to (3. 3) for a general developable is

$$y = x^2 + px^5/10 + p'x^6/60 + \cdots. \tag{6.2}$$

To see that this is true we need only remark that with the coördinate system chosen as it is the equation of the cone cannot contain the variable z and must be the same as the equation of the director curve $y(v)$ in the plane $z = 0$. The expansion (6.2) is well known to be the latter equation (see Lane [2, p. 57]) when the curve $y(v)$ satisfies the differential equation (6.1) and the coördinate system is selected as it is.

Employing this expansion we readily find that there exists a two-parameter family of cubic surfaces with sixth-order contact with S at P and that all these cubic surfaces are actually cubic cones with the same vertex as S. We conclude that the theory of contact of a cubic surface with a cone is coextensive with the theory of contact of a plane cubic curve with a plane curve, at least for orders of contact greater than five. It follows from this latter theory that there exists a one-parameter family of cubic surfaces (actually cubic cones) with seventh-order contact, and that ordinarily there is a unique osculating cubic cone with eighth-order contact. If at every point of the cone the osculating cubic cone has ninth-order contact, then the original cone is itself a cubic cone.

II. Nondevelopable Ruled Surfaces.

7. The differential equations. As is well known, the parametric vector equation of a nondevelopable ruled surface S can be written in the form

$$x(u, v) = y(v) + uz(v),$$

where $y(v)$ and $z(v)$ define director curves on S which satisfy differential equations of the form (see Lane [2, p. 165])

$$y'' = -\tfrac{1}{2}B(v)y + C(v)z, \qquad z'' = -A(v)y + \tfrac{1}{2}B(v)z. \tag{7.1}$$

It follows that $x(u, v)$ satisfies the completely integrable system of partial differential equations

$$x_{uu} = 0, \qquad x_{vv} = -\tfrac{1}{2}\gamma_u x + \gamma x_u, \tag{7.2}$$

where γ is defined by

$$\gamma = Au^2 + Bu + C. \tag{7.3}$$

Conversely, if (7.2) is a completely integrable system, then γ is of the form (7.3) and $x = y + uz$, where y and z satisfy the differential equations (7.1).

If $\gamma \equiv 0$, the integral surfaces of (7.3) are quadrics, while if $\gamma \not\equiv 0$, the conditional equation $\gamma = 0$ defines the flecnode curves on S. We shall hereafter suppose that $\gamma \not\equiv 0$ on the region of the surface in which we are interested.

The most general transformation of asymptotic parameters u, v and proportionality factor which leaves invariant the form of (7.2) is given by

$$u^* = U(u), \qquad v^* = V(v), \qquad x = \tau x^*, \qquad \tau U'V' \neq 0, \tag{7.4}$$

where τ, U and V are defined by

$$\begin{aligned} \tau = (au+b)(cv+d), \qquad U &= (eu+f)/(au+b), \\ V &= (gv+h)/(cv+d), \end{aligned} \tag{7.5}$$

the quantities a, b, c, d, e, f, g and h being constants such that

$$(eb-af)(gd-hc) \neq 0.$$

Under this transformation γ is carried into

$$\gamma^* = U'\gamma/V'^2 = A^*u^{*2} + B^*u^* + C^*,$$

where A^*, B^* and C^* are defined by

$$\begin{aligned} A^* &= (gd-hc)^2(Ab^2 - Bba + Ca^2)/(eb-af)(cv^*-g)^4, \\ B^* &= -(gd-hc)^2(2Abf - Baf - Bbe + 2Cae)/(eb-af)(cv^*-g)^4, \\ C^* &= (gd-hc)^2(Af^2 - Bfe + Ce^2)/(eb-af)(cv^*-g)^4. \end{aligned} \tag{7.6}$$

8. The power series expansions. Let P_x be an ordinary point on S. Since S is not developable, the points x, x_u, x_v, x_{uv} may be chosen as the vertices of a local tetrahedron of reference with unit point so selected that a point $X = x_1x + x_2x_u + x_3x_v + x_4x_{uv}$ will have local homogeneous coördinates proportional to (x_1, x_2, x_3, x_4). If $X = x(u+\Delta u, v+\Delta v)$ is a point on S near x, we can employ a Taylor series and the differential equations (7.2) in a familiar fashion to express each of the local coördinates of X as a power series in Δu and Δv. These series, to as many terms as will be needed in the subsequent discussion, are as follows:

$$\begin{aligned} x_1 = {} & 1 - \tfrac{1}{4}\gamma_u\Delta v^2 - \tfrac{1}{4}\gamma_{uu}\Delta u\Delta v^2 - \tfrac{1}{12}\gamma_{uv}\Delta v^3 - \tfrac{1}{12}\gamma_{uuv}\Delta u\Delta v^3 \\ & + \tfrac{1}{96}(\gamma_u^2 - 2\gamma\gamma_{uu} - 2\gamma_{uvv})\Delta v^4 + \cdots, \\ x_2 = {} & \Delta u + \tfrac{1}{2}\gamma\Delta v^2 + \tfrac{1}{4}\gamma_u\Delta u\Delta v^2 + \tfrac{1}{6}\gamma_v\Delta v^3 + \tfrac{1}{12}\gamma_{uv}\Delta u\Delta v^3 \\ & + \tfrac{1}{24}\gamma_{vv}\Delta v^4 + \tfrac{1}{96}(\gamma_u^2 - 2\gamma\gamma_{uu} + 2\gamma_{uvv})\Delta u\Delta v^4 \\ & + \tfrac{1}{120}(\gamma_u\gamma_v - \gamma\gamma_{uv} + \gamma_{vvv})\Delta v^5 + \cdots, \end{aligned}$$

$$
\begin{aligned}
(8.1)\qquad x_3 &= \Delta v - \tfrac{1}{12}\gamma_u \Delta v^3 - \tfrac{1}{12}\gamma_{uu}\Delta u \Delta v^3 - \tfrac{1}{24}\gamma_{uv}\Delta v^4 \\
&\quad - \tfrac{1}{24}\gamma_{uuv}\Delta u \Delta v^4 + \tfrac{1}{480}(\gamma_u^2 - 2\gamma\gamma_{uu} - 6\gamma_{uvv})\Delta v^5 + \cdots, \\
x_4 &= \Delta u \Delta v + \tfrac{1}{6}\gamma \Delta v^3 + \tfrac{1}{12}\gamma_u \Delta u \Delta v^3 + \tfrac{1}{12}\gamma_v \Delta v^4 + \tfrac{1}{24}\gamma_{uv}\Delta u \Delta v^4 \\
&\quad + \tfrac{1}{40}\gamma_{vv}\Delta v^5 + \tfrac{1}{480}(\gamma_u^2 - 2\gamma\gamma_{uu} + 6\gamma_{uvv})\Delta u \Delta v^5 \\
&\quad + \tfrac{1}{720}(\gamma_u\gamma_v - \gamma\gamma_{uv} + 4\gamma_{vvv})\Delta v^6 + \cdots.
\end{aligned}
$$

If nonhomogeneous coördinates are introduced by placing

$$x = x_2/x_1, \qquad y = x_3/x_1, \qquad z = x_4/x_1,$$

we find that x, y and z may be expressed in power series as follows:

$$
\begin{aligned}
x &= \Delta u + \tfrac{1}{2}\gamma \Delta v^2 + \tfrac{1}{2}\gamma_u \Delta u \Delta v^2 + \tfrac{1}{6}\gamma_v \Delta v^3 + \tfrac{1}{4}\gamma_{uu}\Delta u^2 \Delta v^2 \\
&\quad + \tfrac{1}{6}\gamma_{uv}\Delta u \Delta v^3 + \tfrac{1}{24}(\gamma_{vv} + 3\gamma\gamma_u)\Delta v^4 + \tfrac{1}{12}\gamma_{uuv}\Delta u^2 \Delta v^3 \\
&\quad + \tfrac{1}{24}(\gamma_{uvv} + 3\gamma\gamma_{uu} + 3\gamma_u^2)\Delta u \Delta v^4 \\
&\quad + \tfrac{1}{120}(\gamma_{vvv} + 4\gamma\gamma_{uv} + 6\gamma_u\gamma_v)\Delta v^5 + \cdots, \\
(8.2)\qquad y &= \Delta v + \tfrac{1}{6}\gamma_u \Delta v^3 + \tfrac{1}{6}\gamma_{uu}\Delta u \Delta v^3 + \tfrac{1}{24}\gamma_{uv}\Delta v^4 + \tfrac{1}{24}\gamma_{uuv}\Delta u \Delta v^4 \\
&\quad + \tfrac{1}{120}(4\gamma_u^2 + 2\gamma\gamma_{uu} + \gamma_{uvv})\Delta v^5 + \cdots, \\
z &= \Delta u \Delta v + \tfrac{1}{6}\gamma \Delta v^3 + \tfrac{1}{3}\gamma_u \Delta u \Delta v^3 + \tfrac{1}{12}\gamma_v \Delta v^4 + \tfrac{1}{4}\gamma_{uu}\Delta u^2 \Delta v^3 \\
&\quad + \tfrac{1}{8}\gamma_{uv}\Delta u \Delta v^4 + \tfrac{1}{120}(3\gamma_{vv} + 5\gamma\gamma_u)\Delta v^5 + \tfrac{1}{12}\gamma_{uuv}\Delta u^2 \Delta v^4 \\
&\quad + \tfrac{1}{120}(9\gamma_u^2 + 7\gamma\gamma_{uu} + 4\gamma_{uvv})\Delta u \Delta v^5 \\
&\quad + \tfrac{1}{720}(16\gamma_u\gamma_v + 9\gamma\gamma_{uv} + 4\gamma_{vvv})\Delta v^6 + \cdots.
\end{aligned}
$$

It is now possible to calculate an expansion for z as a power series in x and y. Thus we obtain

$$
\begin{aligned}
(8.3)\qquad z &= xy - \tfrac{1}{3}\gamma y^3 - \tfrac{1}{3}\gamma_u x y^3 - \tfrac{1}{12}\gamma_v y^4 - \tfrac{1}{6}\gamma_{uu} x^2 y^3 - \tfrac{1}{12}\gamma_{uv} x y^4 \\
&\quad + \tfrac{1}{60}(10\gamma\gamma_u - \gamma_{vv}) y^5 - \tfrac{1}{24}\gamma_{uuv} x^2 y^4 \\
&\quad + \tfrac{1}{60}(10\gamma\gamma_{uu} + 10\gamma_u^2 - \gamma_{uvv}) x y^5 \\
&\quad + \tfrac{1}{360}(20\gamma_u\gamma_v + 15\gamma\gamma_{uv} - \gamma_{vvv}) y^6 + \cdots.
\end{aligned}
$$

The expansion (8.3) *for one nonhomogeneous coördinate z as a power series in the other two nonhomogeneous coördinates x and y of a point represents an integral surface of equations* (7.2) *in a sufficiently small neighborhood of an ordinary point x whose local nonhomogeneous coördinates are* 0, 0, 0, *the vertices of the local tetrahedron of reference being the points x, x_u, x_v, x_{uv} and the unit point being suitably chosen.* We remark that all the terms in the series (8.1), (8.2) and (8.3) except those of the highest degree can be obtained from more general series to be found in the treatise of Lane [2, pp. 127-129]. In fact we merely have to put $p = \theta_u = \theta_v = \beta = 0$, $q = -\tfrac{1}{2}\gamma_u$ in the series (V: 4.3), (V: 4.6) and (V: 4.7) of his book in order to obtain these terms.

9. Contact of a cubic surface with a nondevelopable ruled surface. We now proceed to the proof of Theorem II. We write down the equation of the most general cubic surface and demand that it be satisfied by the power series (8. 3) identically in x and y as far as terms of the fifth degree. Thus we find that the equation of any cubic surface having fifth-order contact with S may be written as

$$(9.1)\quad (1/3)\gamma y^3 + Dz^3 + (z - xy)(Px - \tfrac{1}{4}(\gamma_v/\gamma)y + Mz + 1) + yz(Iy + Jz) = 0,$$

where D is an arbitrary parameter and P, M, I and J are defined by the following formulas:

$$\begin{aligned} P &= (15\gamma_v^2 + 40\gamma^2\gamma_u - 12\gamma\gamma_{vv})/80\gamma^3, \\ M &= (40\gamma^2\gamma_u\gamma_v + 12\gamma\gamma_v\gamma_{vv} - 80\gamma^3\gamma_{uv} - 15\gamma_v^3)/320\gamma^4, \\ I &= (5\gamma_v^2 + 40\gamma^2\gamma_u - 4\gamma\gamma_{vv})/80\gamma^2, \\ J &= (15\gamma_u\gamma_v^2 + 40\gamma^2\gamma_u^2 - 12\gamma\gamma_u\gamma_{vv} + 40\gamma^3\gamma_{uu})/240\gamma^3. \end{aligned}$$

If we examine the terms of the sixth degree we see that three of the directions of the sextuple point which the curve of intersection of the surface S and the cubic surface (9. 1) has at P_x coincide in the direction $dv = 0$, i. e., the direction of the generator through P_x, and the other three satisfy the equation

$$(D - P\gamma_{uu}/6)du^3 + \mathcal{A}du^2dv + \mathcal{B}\,dudv^2 + \mathcal{C}\,dv^3 = 0,$$

where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are defined by the following formulas:

$$\begin{aligned} \mathcal{A} &= [(15\gamma_v^2 - 12\gamma\gamma_{vv} - 40\gamma^2\gamma_u)(\gamma_u\gamma_v - \gamma\gamma_{uv}) \\ (9.2)\qquad &\quad + 40\gamma^3(\gamma_u\gamma_v - \gamma\gamma_{uv})_u]/960\gamma^4, \\ \mathcal{B} &= [1600\gamma^4(2\gamma\gamma_{uu} - \gamma_u^2) + 960\gamma^3(\gamma_u\gamma_v - \gamma\gamma_{uv})_v \\ (9.3)\qquad &\quad - 2400\gamma^2\gamma_v(\gamma_u\gamma_v - \gamma\gamma_{uv}) + (12\gamma\gamma_{vv} - 15\gamma_v^2)^2]/57600\gamma^4, \\ \mathcal{C} &= [-20\gamma^2(\gamma_u\gamma_v - \gamma\gamma_{uv}) - 4\gamma^2\gamma_{vvv} + 18\gamma\gamma_v\gamma_{vv} - 15\gamma_v^3]/1440\gamma^2. \end{aligned}$$

We conclude that there exists a unique cubic surface with fifth-order contact and intersecting S in a curve with a sextuple point, four of whose directions coincide in the direction of the generator. The remaining two satisfy

$$\mathcal{A}du^2 + \mathcal{B}\,dudv + \mathcal{C}\,dv^2 = 0.$$

It is obvious that there exists a cubic surface having sixth-order contact with S if and only if $\mathcal{A} = \mathcal{B} = \mathcal{C} = 0$ and that then it is unique. In particular this will be true at every point of S if S is itself a cubic surface. To complete the proof of Theorem II, it is sufficient to prove the converse of this last statement.

Suppose now that $\mathfrak{A} \equiv \mathfrak{B} \equiv \mathfrak{C} \equiv 0$. From (7.3) we find that

$$\gamma_u\gamma_v - \gamma\gamma_{uv} = \pi u^2 - 2\rho u + \sigma,$$

where

$$\pi = AB' - A'B, \qquad \rho = A'C - AC', \qquad \sigma = BC' - B'C.$$

Then we find that $960\gamma^4\mathfrak{A}$ is a polynomial of the sixth degree in u whose leading coefficient is $(\ldots)\pi + 80A^3\rho$ and in which the coefficient of u^5 is $(\ldots)\pi + (\ldots)\rho - 80A^3\sigma$. Since $\mathfrak{A} \equiv 0$, these coefficients must vanish. If $\pi \equiv 0$, then either $A \equiv 0$ or $\rho \equiv 0$. In either case $\rho \equiv A'C - AC' \equiv 0$. Then either $A \equiv 0$ or $\sigma \equiv 0$. If $A \equiv 0$, we find from (9.2) that $960\gamma^4\mathfrak{A}$ is a quadratic polynomial in u whose coefficients are, respectively, given by

$$\begin{aligned} L_1 &= (-40B^3 - 12BB'' + 15B'^2)\sigma, \\ M_1 &= -(80B^2C + 12BC'' + 12B''C - 30B'C')\sigma, \\ N_1 &= (-40BC^2 - 12CC'' + 15C'^2)\sigma. \end{aligned}$$

Since $\mathfrak{A} \equiv 0$, these coefficients must vanish. Therefore

$$0 = C^2L_1 - BCM_1 + B^2N_1 = 15\sigma^3.$$

We conclude that $\pi \equiv 0$ implies $\rho \equiv \sigma \equiv 0$, $\gamma_u\gamma_v - \gamma\gamma_{uv} \equiv 0$. If the last of these equations holds, it is easy to show that

$$(9.5) \qquad \gamma = f(u)g(v) = (lu^2 + mu + n)g(v),$$

where l, m and n are constants not all zero and $g(v) \neq 0$.

The equation $\mathfrak{A} = 0$ may be regarded as a linear homogeneous differential equation satisfied by $\eta(u, v) \equiv \gamma_u\gamma_v - \gamma\gamma_{uv}$ with independent variable u. Uniqueness theorems for such differential equations imply that if there exists a point (u_0, v_0) for which $\eta(u_0, v_0) = 0$, then $\eta(u, v_0) \equiv 0$ in u. Suppose that there is a value v_0 for which $\pi(v_0) \neq 0$. We put

$$u_0 = \{\rho(v_0) + [\rho^2(v_0) - \pi(v_0)\sigma(v_0)]^{\frac{1}{2}}\}/\pi(v_0).$$

Then $\eta(u, v_0) \equiv 0$ and this is a contradiction. Hence $\pi \equiv 0$, so that $\rho \equiv \sigma \equiv 0$ by our proof above. Thus we have proved

LEMMA 9.1. *The equations* $\gamma_{uuu} = 0$, $\mathfrak{A} = 0$ *hold if and only if* γ *is given by* (9.5).

If $\mathfrak{A} \equiv 0$ we can reduce the equation $\mathfrak{B} \equiv 0$ by means of (9.5) to the form

$$(9.6) \qquad \zeta \equiv 12gg'' - 15g'^2 + 40\delta g^3 = 0, \qquad \delta^2 = m^2 - 4ln,$$

and we can reduce the equation $\mathcal{C} \equiv 0$ to the form

$$(9.7) \qquad \theta \equiv -4g^2g''' + 18gg'g'' - 15g'^3 = 8g^{7/2}(g^{-\frac{1}{2}})''' = 0.$$

Then the identity $g\zeta' - 3g'\zeta \equiv -3\theta$ proves the following

LEMMA 9.2. *If* $\gamma_{uuu} = \mathcal{A} = \mathcal{B} = 0$, *then* $\mathcal{C} = 0$.

If $\mathcal{C} \equiv 0$, it follows from (9.7) that g must have the form

$$g(v) = 1/(\lambda v^2 + \mu v + \nu)^2,$$

where λ, μ and ν are constants not all zero. Such a function will satisfy (9.6) if and only if

$$(9.8) \qquad 3\epsilon^2 = -10\delta, \qquad \epsilon^2 = \mu^2 - 4\lambda\nu, \qquad \delta^2 = m^2 - 4ln.$$

Therefore, γ satisfies $\gamma_{uuu} = \mathcal{A} = \mathcal{B} = \mathcal{C} = 0$ if and only if

$$(9.9) \qquad \gamma = (lu^2 + mu + n)/(\lambda v^2 + \mu v + \nu)^2,$$

where the constants satisfy (9.8).

We are now in a position to complete the proof of Theorem II by showing that any integral surface of (7.2) with γ defined by (9.9) is necessarily a cubic surface.

If $\delta = 0$, so that $\epsilon = 0$, we make the transformation (7.4), (7.5) selecting a, b so that $lb^2 - mba + na^2 = 0$. Then $A^* = 0$. Since the equation $B^2 - 4AC = 0$ is invariant, it follows that $B^* = 0$. We choose e, f so that $be - af = 1$. Since $\epsilon = 0$, if we choose c, d so that $\lambda d^2 - \mu dc + \nu c^2 = 0$, then also $2\lambda dh - \mu dg - \mu ch + 2\nu cg = 0$. Then we find that

$$C^* = (gd - hc)^2(lf^2 - mfe + ne^2)/(\lambda h^2 - \mu gh + \nu g^2)^2 \neq 0.$$

Since $d = \xi\nu^{\frac{1}{2}}$, $c = \xi\lambda^{\frac{1}{2}}$, where ξ is arbitrary, we can make $C^* = 3$ by picking $\xi = 3^{\frac{1}{2}}(lf^2 - mfe + ne^2)^{-\frac{1}{2}}$, and the choice of g, h is arbitrary except that $gd - hc \neq 0$. Thus we see that we may assume that the equations (7.2) have the form

$$x_{uu} = 0, \qquad x_{vv} = 3x_u.$$

Four linearly independent solutions of these equations are

$$x_1 = 1, \qquad x_2 = v, \qquad x_3 = u + 3v^2/2, \qquad x_4 = uv + \tfrac{1}{2}v^3.$$

These are the parametric equations of a cubic surface known as Cayley's cubic scroll (see Lane [2, p. 136] or Wilczynski [3, p. 145]). Its implicit equation is $x_1(x_1x_4 - x_2x_3) + x_2{}^3 = 0$.

If $\delta \neq 0$, so that $\epsilon \neq 0$, we make the transformation (7.4), (7.5) selecting a, b, c, d, e, f, g and h so that

$$\begin{gathered} lb^2 - mba + na^2 = lf^2 - mfe + ne^2 = 0, \\ \lambda d^2 - \mu dc + \nu c^2 = \lambda h^2 - \mu hg + \nu g^2 = 0, \\ (be - af)(gd - hc) \neq 0. \end{gathered} \tag{9.10}$$

Then $A^* = C^* = 0$ and B^* is given by

$$B^* = -\frac{(gd - hc)^2(2lbf - maf - mbe + 2nae)}{v^{*2}(be - af)(2\lambda dh - \mu dg - \mu ch + 2\nu cg)^2}.$$

Making use of equations (9.10) we find that $B^* = \delta/\epsilon^2 v^{*2}$. Then from (9.8) we conclude that $B^* = -3/10v^{*2}$. Therefore, we may assume that the equations (7.2) have the form

$$x_{uu} = 0, \qquad x_{vv} = -3x/20v^2 + 3ux_u/10v^2.$$

Four linearly independent solutions of these equations are given by

$$x_1 = v^{1-r}, \qquad x_2 = v^r, \qquad x_3 = v^{1-s}u, \qquad x_4 = v^s u,$$

where $r = \frac{1}{2} + 10^{-\frac{1}{2}}$, $s = \frac{1}{2} + 2 \cdot 10^{-\frac{1}{2}}$. This is also a cubic surface and its implicit equation is $x_1{}^2 x_4 - x_2{}^2 x_3 = 0$. For a discussion of this cubic surface, see Wilczynski [3, p. 145].

These remarks complete the proof of Theorem II and also prove the known

COROLLARY. *Besides the cubic cones, there are only two projectively inequivalent cubic ruled surfaces.*

10. Geometrical interpretations of the invariants $\mathcal{A}$ and $\mathcal{B}$. We have shown that $\mathcal{A} = 0$ if, and only if, $\gamma_u \gamma_v - \gamma \gamma_{uv} = 0$. It then follows from some results stated in Lane [2, pp. 165-166] that $\mathcal{A} = 0$ if, and only if, the flecnode curves are asymptotic curves, or if, and only if, S possesses two (distinct or coincident) straight line director curves. These are the flecnode curves of the preceding characterization. Moreover, $\mathcal{A} = 0$ if, and only if, S belongs to a linear congruence and the directrices of the congruence are the flecnode curves above. Finally, $\mathcal{A} = 0$ if, and only if, the curved asymptotics belong to linear complexes.

To interpret $\mathcal{B}$ we consider the curve of intersection of S with its tangent plane at the point P_x. The intersection has two branches, one of which is the generator $y = z = 0$ and the other of which may be obtained by setting $z = 0$ in the power series (8.3) and solving for x as a power series in y. Thus we find that the equations of this branch are

$$x = \tfrac{1}{3}\gamma y^2 + \tfrac{1}{12}\gamma_v y^3 + \tfrac{1}{180}(3\gamma_{vv} - 10\gamma\gamma_u)y^4 \\ - \tfrac{1}{360}(10\gamma_u\gamma_v + 5\gamma\gamma_{uv} - \gamma_{vvv})y^5 + \cdots, \quad z = 0. \tag{10.1}$$

The equation of the osculating conic of the curve (10.1) is found to be

$$z = Px^2 + x - \tfrac{1}{3}\gamma y^2 - \tfrac{1}{4}(\gamma_v/\gamma)xy = 0. \tag{10.2}$$

We see from (9.1) that this conic may also be given the following characterization. Any cubic surface with fifth-order contact with S intersects the tangent plane in a cubic curve which decomposes into the generator $y = 0$ and the conic (10.2). Our desired interpretation of $\mathcal{B}$ is supplied by the fact that the conic (10.2) hyperosculates the curve (10.1) if, and only if, $\mathcal{B} = 0$.

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY

1. Lane, E. P., "The contact of a cubic surface with an analytic surface," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 471-480.
2. Lane, E. P., *A Treatise on Projective Differential Geometry*, University of Chicago Press, Chicago, 1942.
3. Wilczynski, E. J., *Projective Differential Geometry of Curves and Ruled Surfaces*, Teubner, Leipzig, 1906.

A LIPSCHITZ CONDITION PRESERVING EXTENSION FOR A VECTOR FUNCTION.*

By F. A. Valentine.

1. **Introduction.** The existence of an extension of the range of definition of a function $f(x)$, defined on a metric space $\mathfrak{M}$ to a metric space $\mathfrak{M}'$, so as to preserve a Lipschitz condition (1), depends upon $\mathfrak{M}$ and $\mathfrak{M}'$. In a previous paper [5], the author has established the extension when $\mathfrak{M}$ and $\mathfrak{M}'$ are both Euclidean planes. Both McShane [3], and the author [5], have established, in different ways, the extension when $\mathfrak{M}$ is a metric space, and when $\mathfrak{M}'$ is the one-dimensional real space. Zorn [6], has shown that if $\mathfrak{M}$ and $\mathfrak{M}'$ are a large class of Minkowski spaces, the extension, in general, does not exist. For another example, the result of a discussion with my colleague W. T. Puckett, see Valentine, [5]. *In this paper it is shown that the extension exists when each $\mathfrak{M}$ and $\mathfrak{M}'$ is,* (1) *the n-dimensional Euclidean space*; (2) *the surface of the n-dimensional Euclidean sphere*; or (3) *the general Hilbert space.* The author wishes to take this opportunity to express again his gratitude to his colleagues Max Zorn and W. T. Puckett who have made helpful suggestions concerning these topics.

The following notation is used. A sphere S_i in $\mathfrak{M}$ with radius r_i and center x_i is the set of points x for which $\| x, x_i \| \leqq r_i$. Any notation in $\mathfrak{M}'$ corresponding to one in $\mathfrak{M}$ will be distinguished by means of a prime. The extension of $f(x)$ so as to preserve a Lipschitz condition

$$\| f(x_1), f(x_2) \|' \leqq \| x_1, x_2 \| \tag{1}$$

necessarily depends upon the following property.

Property E. *Consider two sets of spheres M and M' contained in the metric spaces $\mathfrak{M}$ and $\mathfrak{M}'$ respectively. Suppose that to each sphere $S_i \varepsilon M$, having center x_i and radius r_i, there corresponds a sphere $S'_i \varepsilon M'$, having center x'_i and radius r'_i. Furthermore suppose that*

$$r_i = r'_i, \qquad \| x'_i, x'_j \|' \leqq \| x_i, x_j \|, \tag{2}$$

---

* Received July 28, 1943; Presented to the American Mathematical Society, March 17, 1943.

for all corresponding spheres S_i *and* S'_i, *and for all corresponding pairs* (S_i, S_j) *and* (S'_i, S'_j).

The spaces $\mathfrak{M}$ *and* $\mathfrak{M}'$ *are said to have the extensibility property E if under the above assumptions the condition*

$$\prod_M S_i \neq 0 \tag{3}$$

implies

$$\prod_{M'} S'_i \neq 0. \tag{4}$$

In establishing the validity of Property E for the case $\mathfrak{M} = \mathfrak{M}' = \mathfrak{R}_n$, the n-dimensional Euclidean space, it is first proved for the range $(i = 1, \cdots, n+1)$. The property E is then proved for arbitrary sets M and M' by means of a theorem of Helly, [2]. An advantage of the above procedure lies in the fact that property E is a *necessary* as well as *sufficient* condition for the extensibility of $f(x)$ so as to be Lipschitz preserving.

2. Property E in $\mathfrak{R}_n$.

THEOREM 1: *Property* E *holds when each* $\mathfrak{M}$ *and* $\mathfrak{M}'$ *is the n-dimensional Euclidean space* $\mathfrak{R}_n$.

Proof. Let S_i $(i = 1, \cdots, n+1)$ be *any* set of $n+1$ spheres in the set M. We shall first prove that condition (3) implies condition (4) for $(i = 1, \cdots, n+1)$. Let $\Delta(x'_1, \cdots, x'_{n+1})$ be the Euclidean simplex (degenerate or non-degenerate) determined by the centers x'_i of S'_i $(i = 1, \cdots, n+1)$. *Suppose that* $\Delta(x'_1, \cdots, x'_{n+1})$ *is not covered by the sets* S'_i. Then choose x and x' so that

$$x \,\varepsilon \prod_{i=1}^{n+1} S_i, \qquad x' \,\varepsilon\, \Delta(x'_1, \cdots, x'_{n+1}) - \sum_{i=1}^{n+1} S'_i, \tag{5}$$

and let R_i and R'_i be the n-dimensional vectors drawn from x and x' to x_i and x'_i respectively. Conditions (5) and (2) imply that

$$R'_i \cdot R'_i > R_i \cdot R_i, \qquad (i \text{ not summed}), \tag{6}$$

$$(R'_i - R'_j) \cdot (R'_i - R'_j) \leqq (R_i - R_j) \cdot (R_i - R_j), \tag{7}$$

where the products involved are vector dot products. Conditions (6) and (7) yield the inequalities

$$R'_i \cdot R'_j > R_i \cdot R_j. \tag{8}$$

Since the point x' is contained in $\Delta(x'_1, \cdots, x'_{n+1})$, there exist constants a'_i $(i = 1, \cdots, n+1)$ such that

$$(9) \qquad a'_i \geqq 0, \qquad \sum_{i=1}^{n+1} a'_i \neq 0, \qquad a'_i R'_i = 0, \qquad (i \text{ summed}).$$

Multiplying (8) by $a'_i a'_j$, summing on i and j, one obtains

$$(10) \qquad (a'_i R'_i) \cdot (a'_j R'_j) > (a'_i R_i) \cdot (a'_j R_j),$$

which by (9) implies that

$$| a'_i R_i |^2 < 0, \qquad (i \text{ summed}),$$

which is impossible. Hence $\Delta(x'_1, \cdots, x'_{n+1})$ is covered by the sets S'_i. Furthermore conditions (3) imply that

$$(11) \qquad S_i \cdot S_j \cdot B(x_j, x_j) \neq 0,$$

where $B(x_i, x_j)$ is the side of the simplex Δ joining x_i and x_j. Conditions (2) and (11) then yield the conditions

$$(12) \qquad S'_i \cdot S'_j \cdot B(x'_i, x'_j) \neq 0.$$

By a theorem of Knaster, Kuratowski and Mazurkiewicz,[1] conditions (12) and the fact that $\Delta(x'_1, \cdots, x'_{n+1})$ is covered by the sets S'_i, imply that $\prod_{i=1}^{n+1} S'_i \neq 0$.

Since we have shown that *each* set of $n+1$ of the spheres S'_i have a point in common, *it follows by a theorem of Helly,*[2] *that all the spheres* S'_i *have a point in common.* Thus Theorem 1 is proved.

3. Property E in $\mathcal{K}_n$. Let $\mathcal{K}_n$ be the surface of the $(n+1)$-dimensional Euclidean sphere E_{n+1} in $\mathcal{R}_{n+1}$. Although Lemmas 1-3 are used to establish property E for $\mathcal{K}_n$, they are of independent geometric interest.

LEMMA 1. *Let* $\mathfrak{M}$ *and* $\mathfrak{M}'$ *be two n-dimensional spherical surfaces* $\mathcal{K}_n$ *and* $\mathcal{K}'_n$ *belonging to two Euclidean spheres* E_{n+1} *and* E'_{n+1}, *respectively,*

---

[1] See Alexandroff and Hopf, [1, p. 377]. The theorem states: *If the closed sets* $A_0, \cdots, A_n$ *cover the simplex T, and if for each side* $a_{i_0}, \cdots, a_{i_r}$ *of T, we have*

$$a_{i_0}, \cdots, a_{i_r} \subset A_{i_0} + \cdots + A_{i_r},$$

then $A_0 \cdot A_1 \cdot \ldots \cdot A_n \neq 0$.

[2] See [2] or [1, p. 297]. The theorem states: *If each* $n+1$ *sets of a family of closed bounded convex sets of the n-dimensional Euclidean space intersect, then there is a point in common to all the sets.*

of equal radii; if in the statement of property E, $i = 1, \cdots, n+1$, *then conditions* (2) *and* (3) *imply conditions* (4).

Proof. Since $i = 1, \cdots, n+1$, *the set of centers x_i of S_i must lie on the surface of a closed hemisphere H_{n+1} of the sphere E_{n+1}.* A corresponding statement holds for the points x'_i. It should be remembered that S_i and S'_i are here n-dimensional spheres in $\mathcal{K}_n$ and $\mathcal{K}'_n$ respectively. We *first assume* that x'_i are interior points of the surface of H'_{n+1}. Denote the surface simplex (degenerate or non-degenerate) on H_{n+1}, determined by the points x_i, by $\sigma(x_1, \cdots, x_{n+1})$. Attach a corresponding meaning to $\sigma(x'_1, \cdots, x'_{n+1})$. Condition (3) implies that

$$S_i \cdot S_j \cdot B(x_i, x_j) \neq 0, \qquad (i, j = 1, \cdots, n+1), \tag{13}$$

where $B(x_i, x_j)$ is the side of the simplex σ joining x_i and x_j.

By virtue of conditions (2) and (13) we have $\| x'_i, x'_j \| \leqq r'_i + r'_j$, so that

$$S'_i \cdot S'_j \cdot B(x'_i, x'_j) \neq 0. \tag{14}$$

Choose the point x so that

$$x \,\varepsilon \prod_{i=1}^{n+1} S_i \neq 0. \tag{15}$$

Suppose that the simplex $\sigma(x'_1, \cdots, x'_{n+1})$ is *not* covered by the spheres S'_i. Hence choose $x' \,\varepsilon\, \sigma(x'_1, \cdots, x'_{n+1}) - \sum_{i=1}^{n+1} S'_i$, which implies that

$$\| x', x'_i \|' > \| x, x_i \| = r_i = r'_i. \tag{16}$$

We now show that (16) contradicts the hypotheses (2). Let R_i, R be the $(n+1)$*-dimensional vectors in $\mathcal{R}_{n+1}$ drawn from the center of the sphere E_{n+1} to the points x_i and x, respectively. Attach a corresponding meaning to* R'_i, R'. Conditions (2) imply that

$$(R'_i - R'_j) \cdot (R'_i - R'_j) \leqq (R_i - R_j) \cdot (R_i - R_j) \tag{17}$$

hold, where these are vector dot products. Furthermore conditions (16) yield the inequalities

$$(R' - R'_i) \cdot (R' - R'_i) > (R - R_i) \cdot (R - R_i). \tag{18}$$

Since $| R_i | = | R'_i | = | R | = | R' |$, conditions (17) and (18) imply that

$$R'_i \cdot R'_j \geqq R_i \cdot R_j, \tag{19}$$

$$(i, j = 1, \cdots, n+1).$$

$$R \cdot R_i > R' \cdot R'_i, \tag{20}$$

Since $x' \varepsilon \sigma(x'_1, \cdots, x'_{n+1})$, and since the x'_i are interior points of the surface of the hemisphere H'_{n+1} there exist, by a well-known property of vectors, constants λ'_i such that

$$(21) \qquad R' = \lambda'_i R'_i, \qquad \lambda'_i \geqq 0, \qquad \sum_{i=1}^{n+1} \lambda'_i \neq 0; \qquad (i \text{ summed}).$$

On multiplying both sides of (19) by $\lambda'_i \lambda'_j$, summing on i and j, one obtains

$$(\lambda'_i R'_i) \cdot (\lambda'_j R'_j) \geqq (\lambda'_i R_i) \cdot (\lambda'_j R_j)$$

which, on account of (21), yields

$$(22) \qquad | \lambda'_i R_i | \leqq | R' |, \qquad (i \text{ summed}).$$

Similarly on multiplying both sides of (20) by λ'_i, summing on i, one obtains

$$R \cdot (R_i \lambda'_i) > R' \cdot R',$$

which implies, since $| R' | = | R |$, that

$$(23) \qquad | \lambda'_i R_i | > | R' |, \qquad (i \text{ summed}).$$

Since (22) and (23) are contradictory, the assumption that $\sigma(x'_1, \cdots, x'_{n+1})$ is not covered by the spheres S'_i is false. Then conditions (14) imply, by the theorem of Knaster, Kuratowski and Mazurkiewicz,[3] that $\prod_{i=1}^{n+1} S'_i \neq 0$, and Lemma 1 has been proved when the x'_i are interior points of the surface of H'_{n+1}.

Since $i = 1, \cdots, n+1$, the remaining case is that in which all of the points x'_i are on the boundary of the surface of H'_{n+1}. In this case there exist sequences of points x'_{mi}, such that $\| x'_{mi}, x'_{mj} \|' \leqq \| x_i, x_j \|$, $x'_{mi} \to x'_i$, as $m \to \infty$, and such that x'_{mi} are interior points of the surface of H'_{n+1}. For example if P is the center of the surface of H'_{n+1}, choose x'_{mi} on the great circle joining x'_i and P so that $\| x'_{mi}, x'_i \|' = \epsilon(m)$, where $\epsilon(m) \to 0$, as $m \to \infty$, for $(i = 1, \cdots, n+1)$. Then by the above proof Lemma 1 holds for the points x_i and x'_{mi}, and hence by passage to the limit, for the points x_i and x'_i.

LEMMA 2. *Let M and M' be two sets of points, each set lying on the surface $\mathcal{K}_n$ of an $(n+1)$-dimensional Euclidean sphere E_{n+1} of radius r. Suppose that to each point $x_i \varepsilon M$ there corresponds a point $x'_i \varepsilon M'$.*

---

[3] *Loc. cit.*[1]

Furthermore suppose that there exists in M' *a set of points* x'_i, $(i = 1, \cdots, m+2, 1 \leqq m \leqq n)$ *which determine an* $(m+1)$*-dimensional Euclidean simplex* $\Delta(x'_1, \cdots, x'_{m+2})$ *containing the center* O *of* E_{n+1} *in its interior. If we have*

$$\| x'_i, x'_j \| \leqq \| x_i, x_j \| \tag{24}$$

for all corresponding pairs in M *and* M'*, respectively, then*

$$\| x'_i, x'_j \| = \| x_i, x_j \| \tag{25}$$

for all such pairs.

Proof. Since $\mathcal{K}'_n = \mathcal{K}_n$, we can omit the primes on the metrics entirely. The lemma is proved *first* for $x_1, \cdots, x_{m+2}$. If the strict equality in (24) holds, the lemma is empty. Hence suppose, without loss of generality, that, for example,

$$\| x'_1, x'_2 \| < \| x_1, x_2 \|. \tag{26}$$

Let (x_1, y) be a *diametrically opposite point pair* in $\mathcal{K}_n$ and attach a corresponding meaning to (x'_1, y'). Then by definition

$$\begin{aligned} &\| x_1, x_i \| + \| x_i, y \| = \pi r, \\ &\| x'_1, x'_i \| + \| x'_i, y' \| = \pi r, \end{aligned} \qquad (i = 2, \cdots, m+2). \tag{27}$$

Conditions (24), (26) and (27) imply that

$$\begin{aligned} &\| x'_i, y' \| \geqq \| x_i, y \|, \\ &\| x'_2, y' \| > \| x_2, y \|. \end{aligned} \qquad (i = 3, \cdots, m+2), \tag{28}$$

Let R_i, R'_i, R, R' *be the* $(n+1)$*-dimensional vectors drawn from the center* O *of* E_{n+1} *to the points* x_i, x'_i, y, y' *respectively.* Since the center O of E_{n+1} is interior to the Euclidean simplex $\Delta(x'_1, \cdots, x'_{m+2})$, the point y' is *interior* to the surface simplex $\sigma(x'_2, \cdots, x'_{m+2})$, and $\sigma(x'_2, \cdots, x'_{m+2})$ must be on the interior of the surface of a hemisphere H_{n+1} of E_{n+1}. This last statement implies the existence of constants $\lambda'_i > 0$ $(i = 2, \cdots, m+2)$, such that $R' = \lambda'_i R'_i$, (i summed). Conditions (24) yield conditions (19) with $(i, j = 2, \cdots, m+2)$. Conditions (28) yield the inequalities,

$$\begin{aligned} &R \cdot R_j \geqq R' \cdot R'_j, \\ &R \cdot R_2 > R' \cdot R'_2. \end{aligned} \qquad (j = 3, \cdots, m+2), \tag{29}$$

Conditions (19) and (29) imply, just as in the proof of Lemma 1, the contradictory results (22) and (23) with $(i=2,\cdots,m+2)$. Since in (26), x_1 and x_2 were chosen without restriction, Lemma 2 is proved for $(i=1,\cdots m+2)$.

Secondly, to complete the proof, let x_s be *any* point in M, with $x_s \neq x_i$, $(i=1,\cdots,m+2)$. Since Lemma 2 holds with $(i=1,\cdots,m+2)$, conditions (25) imply that the $(n+1)$-dimensional Euclidean simplex $\Delta(x'_1,\cdots,x'_{m+2})$ can be moved rigidly so that $x'_i=x_i$. Hence move the set M' rigidly in $\mathcal{R}_{n+2}$ so that $x'_i=x_i$, $(i=1,\cdots,m+2)$. *Suppose* that Lemma 2 is false, so that, for example,

$$(30) \qquad \| x'_1, x'_s \| < \| x_1, x_s \|.$$

Since conditions (27) hold with i replaced by s, conditions (30) give the result

$$(31) \qquad \| x'_s, y' \| > \| x_s, y \|.$$

Conditions (24) and (31) imply that

$$(32) \qquad R'_i \cdot R'_s \geqq R_i \cdot R_s, \qquad (i=2,\cdots,m+2),$$

$$(33) \qquad R \cdot R_s > R' \cdot R'_s.$$

Conditions (32) yield the result $(\lambda'_i R'_i) \cdot R'_s \geqq (\lambda'_i R_i) \cdot R_s$, ($i$ summed). Since $x'_i=x_i$, $y=y'$, we have $R'_i=R_i$, $R'=R$, whence by the previous sentence $R'\cdot R'_s \geqq R \cdot R_s$, which contradicts condition (33). Thus Lemma 2 is proved.

LEMMA 3. *Let M and M' be two sets of points, each set lying on the surface $\mathcal{K}_n$ of an $(n+1)$-dimensional Euclidean sphere E_{n+1} of radius r, and suppose that to each point $x_i \varepsilon M$ there corresponds a point $x'_i \varepsilon M'$. If the set of points M' do not all lie on the surface of a hemisphere H_{n+1} of E_{n+1}, and if conditions (24) hold for all corresponding pairs (x_i, x'_i) and (x_j, x'_j) in M and M', then conditions (25) hold for all such pairs.*

Proof. We use the same notation used in the proof of Lemma 2. Choose an *arbitrary* point in M, and without loss of generality denote it by x_1. *Suppose* that y', the point diametrically opposite to x'_1 on E_{n+1}, is such that $y' \varepsilon M'$. Let x_s be an *arbitrary* point in M. Then since

$$\| x_1, x_s \| + \| x_s, y \| = \pi r, \qquad \| x'_1, x'_s \| + \| x'_s, y' \| = \pi r,$$
$$\| x'_1, x'_s \| \leqq \| x_1, x_s \|, \qquad \| x'_s, y' \| \leqq \| x_s, y \|,$$

we must have $\| x'_1, x'_s \| = \| x_1, x_s \|$. Hence conditions (25) hold in the first case. The same proof holds if y' is a limit point of M'.

Secondly, suppose $y' \notin \bar{M}'$, the closure of M'. Let $C(M')$ be the convex hull of M'. Since $M' \notin H_{n+1}$, $C(M')$ contains the center O of E_{n+1} in its *interior*. Hence the line joining x'_1 and O meets when extended the surface of $C(M')$ at a point $P \notin \bar{M}'$. Hence let $\Delta(x'_2, \cdots, x'_{m+2})$ $(0 \leqq m \leqq n)$ be the smallest simplex of the surface of $C(M')$ containing P in its *interior*. Then the simplex $\Delta(x'_1, x'_2, \cdots, x'_{m+2})$ contains O in its interior, and hence satisfies the hypothesis of Lemma 2. Thus (25) holds, and Lemma 3 is proved. By using the closure of M', the above argument yields the

Remark. Suppose that $M' \varepsilon \mathcal{K}_n$, the surface of E_{n+1} of radius r. If each set of $m+2$ $(0 \leqq m \leqq n)$ distinct points of M' lie on a hemisphere of E_{m+1} of radius r, then M' lies on a hemisphere of E_{n+1}.

THEOREM 2. *Property* E *holds when each* $\mathfrak{M}$ *and* $\mathfrak{M}'$ *is the n-dimensional surface* $\mathcal{K}_n$ *of the* $(n+1)$*-dimensional Euclidean sphere* E_{n+1}.

The proof of this theorem falls naturally into two cases.

Case I. Suppose that the centers x'_i of the spheres $S'_i \, \varepsilon \, M'$ all lie on the surface of a closed hemisphere H_{n+1} of E_{n+1}. By Lemma 1, *each set of* $n+1$ *of the sets* $S'_i \cdot H_{n+1}$ *have a point in common.* The projection of the sets $S'_i \cdot H_{n+1}$ on the base plane of the hemisphere H_{n+1} are convex sets, which we denote by C'_i. Since a projection sets up a one-to-one correspondence between points of the surface of H_{n+1} and points of the base plane, each set of $n+1$ of the sets C'_i have a point in common. *By a theorem of Helly*[4] [2], *or* [1, p. 297], *it follows that all the sets* C'_i *have a point in common.* Thus on account of the one-to-one correspondence, $H_{n+1} \cdot \Pi_{M'} S'_i \neq 0$, and Theorem 2 is proved for this case.

Case II. In this case suppose that the centers x'_i do *not* all lie on the surface of a hemisphere of E_{n+1}. Then by Lemma 3, $\| x'_i, x'_j \| = \| x_i, x_j \|$. Hence by a rigid motion of M' in $\mathfrak{R}_{n+2}$ we can make $x'_i = x_i$ for all corresponding pairs S_i and S'_i. Since $r_i = r'_i$, we have $S_i = S'_i$. Thus $\Pi_M S_i \neq 0$ is identical with $\Pi_{M'} S'_i \neq 0$. Consequently Theorem 2 has been proved in all cases.

4. Property E in $\mathfrak{H}$.

Let $\mathfrak{H}$ be a general Hilbert space, the number of dimensions not necessarily being denumerable.

THEOREM 3. *Property* E *holds when each* $\mathfrak{M}$ *and* $\mathfrak{M}'$ *is a Hilbert space* $\mathfrak{H}$.

---

[4] *Loc. cit.*[2]

This theorem is a consequence of the following statement. *If each finite set of the spheres S_i in $\mathfrak{H}$ have a point in common, then they all have a point in common.* To prove this let the spheres S_i be well-ordered, and let the corresponding products $\prod_{i=1}^{\alpha} S_i = T_\alpha$, ($\alpha =$ transfinite ordinal). Since the sets T_α are well-ordered, *suppose that the first* one which is zero is T_β. If β is a limiting ordinal, then since the space $\mathfrak{H}$ is regular, and since the sets T_α are a transfinite bounded monotone decreasing sequence of closed convex sets, a theorem of Smulian [4] states that $T_\beta \neq 0$, contrary to our assumption.[5] The remaining case is that in which the first of the T_α which is zero is T_β, with $\beta = \delta + n$, where δ is a limiting ordinal and n is an integer. In this case reorder the spheres S_i, so that the first set of the sets $\prod_{i=1}^{\alpha} S_i = T_\alpha{}^1$ which is zero is $T_{\beta_1}{}^1$, where $\beta_1 \leqq \delta$. If β_1 is *not* a limiting ordinal, repeat this process so as to obtain a sequence of sets $T_{\beta_\lambda}{}^\lambda$ (β_λ and λ are ordinals), so that if $\lambda > \mu$, $\beta_\lambda < \beta_\mu$. The ordinals β_λ must then have a smallest member, denoted by β_γ. The number β_γ must, by construction, be either a limiting ordinal or an integer. In the former case we have, by the Smulian theorem [4], $T_{\beta_\gamma}{}^\gamma \neq 0$, whereas in the latter case this condition follows by assumption. Since in all cases $T_\alpha{}^\lambda \neq 0$, ($\lambda =$ fixed ordinal, $\alpha =$ any ordinal), the statement above is proved.

In a Hilbert space any $n + 1$ spheres S_i must have intersections in the same n-dimensional Euclidean space $\mathfrak{R}_n$. Since for any $n + 1$ spheres S_i in $\mathfrak{H}$, the n-dimensional Euclidean spheres $S_i \cdot \mathfrak{R}_n$, $S'_i \cdot \mathfrak{R}'_n$ satisfy the hypotheses of Theorem 1, we must have $\prod_{i=1}^{n+1} S'_i \neq 0$. Since n is an arbitrary integer, by the italicized theorem in the preceding paragraph we must have $\Pi_{M'} S'_i \neq 0$. Hence Theorem 3 is proved.

5. The extension.

THEOREM 4. *Let $f(x)$ be a function which maps a set S in a metric space $\mathfrak{M}$ into a set S' in a metric space $\mathfrak{M}'$, so as to preserve the Lipschitz condition. Furthermore suppose that $\mathfrak{M}$ and $\mathfrak{M}'$ have the property* E. *Then $f(x)$ can be extended to any set $T \supset S$ so as to preserve the Lipschitz condition.*

A Lipschitz extension is possible as soon as one establishes the extensibility of $f(x)$ from an *arbitrary set U to any additional point x_0*. To do this here let S_i be spheres in $\mathfrak{M}$ with centers x_i and radii $r_i = \| x_0, x_i \|$, respectively, and let S'_i be spheres in $\mathfrak{M}'$ with centers $x'_i \equiv f(x_i)$ and radii $r'_i = r_i$ respectively. Let x_i range over U and denote the corresponding set of spheres by M. Since we have assumed that property E holds, any point $x'_0 \,\varepsilon\, \Pi_{M'} S'_i \neq 0$ serves as an extension $f(x_0)$, for $\| x'_0, f(x_i) \|' \leqq \| x_0, x_i |$ for all corresponding pairs $[x_i, f(x_i)]$ in U and U'. If the spaces $\mathfrak{M}$ and $\mathfrak{M}'$ are separable, the function $f(x)$ can be extended to a denumerable dense set by ordinary induction, and then to the whole space by passage to the limit. For example, see [5, p. 106]. If $\mathfrak{M}$ and $\mathfrak{M}'$ are not separable, the extension follows by transfinite induction.

---

[5] The Smulian theorem and its applicability here were suggested to me by my colleague Max Zorn. The italicized statement in the above proof also holds when $\mathfrak{H}$ is a regular Banach space and when S_i are bounded, closed, convex sets.

THEOREM 5. *The extension described in Theorem 4 holds for the cases;* (1) $\mathfrak{M} = \mathfrak{M}' = \mathcal{R}_n$; (2) $\mathfrak{M} = \mathfrak{M}' = \mathcal{K}_n$; (3) $\mathfrak{M} = \mathfrak{M}' = \mathfrak{H}$. *Furthermore for cases* (1) *and* (3) *the extension* $[f(x), x \,\varepsilon\, T]$ *can be defined so as to be contained in any closed, convex set* $L \supset S'$.

The first statement follows from Theorem 4, and the fact that $\mathcal{R}_n$, $\mathcal{K}_n$ and $\mathfrak{H}$ have property E.

To prove the concluding statement for the space $\mathcal{R}_n$, let S'_i $(i = 1, \cdots, n+1)$ be any $n+1$ spheres in M'. Then the simplex $\Delta(x'_1, \cdots, x'_{n+1}) \subset L$. Since in the proof of Lemma 1, $\Delta(x'_1, \cdots, x'_{n+1}) \cdot \prod_{i=1}^{n+1} S'_i \neq 0$, we have $L \cdot \prod_{i=1}^{n+1} S'_i \neq 0$. This together with the fact that $\Pi_{M'} S'_i \neq 0$, implies by a theorem of Helly[6] [2], that $L \cdot \Pi_{M'} S'_i \neq 0$. Thus we can choose $x'_0 \,\varepsilon\, L \cdot \Pi_{M'} S'_i$, so that the extension $f(x_0) \,\varepsilon\, L$. By induction, it follows that the total extension can be defined so that $[f(x), x \,\varepsilon\, T] \subset L$.

A corresponding result holds for a Hilbert space $\mathfrak{H}$, since by the preceding reasoning we have $L \cdot \prod_{i=1}^{n+1} S'_i \neq 0$, where n is an *arbitrary* integer. Hence by virtue of the italicized statement in the proof of Theorem 3, we have $L \cdot \Pi_{M'} S'_i \neq 0$, and the rest of the argument is the same as that for $\mathcal{R}_n$.

COROLLARY. *Suppose that the function* $f(z)$ *maps a set* S *in* $\mathcal{R}_n$ *into a set* S' *in* $\mathcal{R}_n$, *and suppose that there exists a constant* $K > 0$ *so that*

$$\| f(z_1), f(z_2) \| \leqq K \, \| z_1, z_2 \| \tag{34}$$

for all pairs z_1 *and* z_2 *in* S. *Then* $f(z)$ *can be extended to any set* $T \supseteq S$ *so as to preserve condition* (34). *This also holds for a Hilbert space* $\mathfrak{H}$. *The concluding statement of Theorem* 5 *also holds.*

---

[6] *Loc. cit.*[2]

This is an immediate consequence of Theorem 4 and of the fact that $k \| z_1, z_2 \| = \| kz_1, kz_2 \|$. In the proof of Theorem 4, let $x_i \equiv kz_i$, $x'_i \equiv f(z_i)$, $x_0 \equiv kz_0$, $r_i = r'_i \equiv \| x_0, x_i \|$. Then by the argument in the proof of Theorem 4, the extension $f(z_0)$ exists. The rest of the proof is the same as that given for Theorems 4 and 5.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES.

BIBLIOGRAPHY.

1. Alexandroff and Hopf, *Topologie*, vol. 1 (1935), Springer, Berlin.
2. E. Helly, *Jber. Deutschen Math. Verein.*, vol. 32 (1923), pp. 175-176.
3. E. J. McShane, "Extension of range of functions," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 837-842.
4. V. Smulian, "On the principle of inclusion in the space of type B," *Rec. Math.* [*Math. Sbornik*] (1939), N. S. 5 (47), pp. 317-328. See also *Math. Rev.*, vol. 1, No. 11 (1940), p. 335.
5. F. A. Valentine, "On the extension of a vector function so as to preserve a Lipschitz condition," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 100-108.
6. Max Zorn, "Contractions in Minkowski spaces." (Unpublished).

NÖRLUND METHODS OF SUMMABILITY THAT INCLUDE THE CESÀRO METHODS OF ALL POSITIVE ORDERS.*

By J. D. HILL.

---

1. Introduction. In a recent paper[1] H. L. Garabedian has established the existence of Hausdorff methods of summability having the unusual property of including the Cesàro methods of all real and positive orders. The object of this paper is to consider the corresponding problem for Nörlund methods. We begin by recalling a few facts on which the sequel depends.

The Nörlund method of summability $[N; \{p_k\}]$ is defined by a triangular matrix (a_{mk}) whose elements are of the form $a_{mk} \equiv p_{m-k}/P_m$ $(k = 0, 1, 2, \cdots, m;$ $m = 0, 1, 2, \cdots)$, where $\{p_k\}$ is a sequence of complex numbers such that $P_m \equiv p_0 + p_1 + \cdots + p_m$ is different from zero for all m. One generally takes $p_0 = 1$ since no loss of generality results. We then have $a_{mm} = P_m^{-1} \neq 0$ for all m, so that the Nörlund transformation

$$(1.1) \qquad t_m = P_m^{-1} \sum_{k=0}^{m} p_{m-k} s_k \qquad (m = 0, 1, 2, \cdots),$$

is *reversible*;[2] that is to say, corresponding to each convergent sequence $\{t_m\}$ there exists a unique (convergent or divergent) sequence $\{s_k\}$ satisfying the equations (1.1).

The Silverman-Toeplitz regularity conditions for $[N; \{p_k\}]$ reduce to

$$(1.2) \qquad \sum_{k=0}^{m} | p_k | = O(P_m) \quad \text{and} \quad p_m = o(P_m) \quad \text{as} \quad m \to \infty.$$

Our results are based on the following general theorem of S. Mazur.[3]

THEOREM OF MAZUR. *Let A with matrix (a_{mk}) be a given reversible*

---

* Received January 17, 1944.

[1] H. L. Garabedian, "Hausdorff methods of summation which include all of the Cesàro methods," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 124-127.

[2] S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 90.

[3] S. Mazur, "Über lineare Limitierungsverfahren," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 599-611. For the theorem in the form stated here in which Mazur's hypothesis of *normality* is replaced by that of *reversibility*, see J. D. Hill, "Some properties of summability," *Duke Mathematical Journal*, vol. 9 (1942), pp. 373-381; in particular, p. 376.

method, and let B with matrix (b_{mk}) *be any method whatsoever. Let the sequences* $\{\xi_k\}$ *and* $\{\xi_k{}^n\}$ *be defined as the (unique) solutions of the equations*

$$\sum_{k=0}^{\infty} a_{mk}\xi_k = 1 \qquad (m = 0, 1, 2, \cdots); \tag{1.3}$$

$$\sum_{k=0}^{\infty} a_{mk}\xi_k{}^n = \delta_m{}^n \qquad (m, n = 0, 1, 2, \cdots); \tag{1.4}$$

where $\delta_m{}^n$ *is the Kronecker symbol. Then in order that B include A the following conditions are necessary and sufficient.*

$$B - \lim \{\xi_k\} \text{ exists and } = 1; \tag{1.5}$$

$$B - \lim \{\xi_k{}^n\} \text{ exists and } = 0 \qquad (n = 0, 1, 2, \cdots); \tag{1.6}$$

$$\sup_r \{Q_r{}^m \equiv \sum_{n=0}^{\infty} | \sum_{k=0}^{r} b_{mk}\xi_k{}^n |\} < \infty \qquad (m = 0, 1, 2, \cdots); \tag{1.7}$$

$$\sup_m \{Q_m \equiv \sum_{n=0}^{\infty} | \sum_{k=0}^{\infty} b_{mk}\xi_k{}^n |\} < \infty. \tag{1.8}$$

2. The condition that $[N; \{p_k\}] \supset (C; \alpha > 0)$. We shall now apply the theorem of Mazur with $A \equiv (C; \alpha > 0)$ and with $B \equiv [N; \{p_k\}]$, a regular Nörlund method. Since $(C; \alpha)'$ is itself a (reversible) regular Nörlund method $[N; \{p_k(\alpha)\}]$ where

$$p_k(\alpha) \equiv \binom{k + \alpha - 1}{\alpha - 1} \text{ and } P_k(\alpha) = \binom{k + \alpha}{\alpha} \qquad (k = 0, 1, 2, \cdots),$$

we see that (a_{mk}) in the preceding theorem becomes a triangular matrix whose elements are $p_{m-k}(\alpha)/P_m(\alpha)$; and that (b_{mk}) is to be replaced by (p_{m-k}/P_m). With these replacements it follows immediately from (1.3) that $\xi_k \equiv \xi_k(\alpha) = 1$ $(k = 0, 1, 2, \cdots)$, and in (1.4) it is known that[4]

$$\xi_k{}^n \equiv \xi_k{}^n(\alpha) = (-1)^{k-n} \binom{n + \alpha}{\alpha}\binom{\alpha}{k - n} \qquad (k, n = 0, 1, 2, \cdots),$$

where the final binomial coefficient is defined as 0 if $k - n < 0$. Since $\lim_k \xi_k(\alpha) = 1$ and $\lim_k \xi_k{}^n(\alpha) = 0$ $(n = 0, 1, 2, \cdots)$ we observe that (1.5) and (1.6) are automatically satisfied on account of the assumed regularity of $[N; \{p_k\}]$. Condition (1.7) is likewise automatically fulfilled since for each m, $Q_r{}^m(\alpha)$ is constant for fixed α and $r > m$.

Our problem is consequently reduced to a study of the conditions under which (1.8) will be satisfied. We find in the present case that

---

[4] J. D. Hill, "On perfect methods of summability," *Duke Mathematical Journal*, vol. 3 (1937), pp. 702-714; in particular, pp. 705-707.

$$(2.1) \qquad Q_m(\alpha) = | P_m |^{-1} \sum_{n=0}^{m} \binom{n+\alpha}{\alpha} \Big| \sum_{j=0}^{m-n} (-1)^j \binom{\alpha}{j} p_{m-n-j} \Big|.$$

To assist in the analysis of this expression we introduce the following power series:

$$(2.2) \qquad p(z) \equiv \sum_{k=0}^{\infty} p_k z^k;$$

$$(2.3) \qquad C_\alpha(z) \equiv (1-z)^\alpha p(z) \equiv \sum_{n=0}^{\infty} c_n(\alpha) z^n,$$

so that

$$(2.4) \qquad c_n(\alpha) = \sum_{j=0}^{n} (-1)^j \binom{\alpha}{j} p_{n-j};$$

$$(2.5) \qquad C^*_\alpha(z) \equiv \sum_{n=0}^{\infty} | c_n(\alpha) | z^n;$$

$$(2.6) \qquad (1-z)^{-\alpha-1} C^*_\alpha(z) \equiv \sum_{m=0}^{\infty} g_m(\alpha) z^m,$$

so that

$$(2.7) \qquad g_m(\alpha) = \sum_{n=0}^{m} \binom{n+\alpha}{\alpha} | c_{m-n}(\alpha) |.$$

Since the series in (2.2) has a positive radius of convergence[5] it is evident that the same is true of the series in (2.3), (2.5), and (2.6). In view of the foregoing definitions we observe that (2.1) reduces to the form $Q_m(\alpha) = | P_m |^{-1} g_m(\alpha)$. The condition (1.8) is therefore equivalent to the condition

$$(2.8) \qquad g_m(\alpha) = O(P_m) \qquad (m \to \infty).$$

We express the results of the preceding discussion in the form of the following theorem.

THEOREM 1. *In order that a given regular* $[N; \{p_k\}]$ *include* $(C; \alpha)$ *for a given* $\alpha > 0$ *it is necessary and sufficient that the condition* (2.8) *be satisfied.*

COROLLARY. *If a regular* $[N; \{p_k\}]$ *is independent of* α *and if* (2.8) *holds for each* $\alpha > 0$, *then* $[N; \{p_k\}]$ *includes* $(C; \alpha)$ *for all positive* α.

To illustrate the applicability of Theorem 1 the reader may easily verify the well known result

$$(2.9) \qquad (C; \beta) \supset (C; \alpha) \qquad (\beta > \alpha > 0).$$

---

[5] See footnote 4, p. 706.

In the next section we derive Theorem 2 as a special case of the preceding corollary, and use it to give an example of a method $[N;\{p_k\}]$ that includes all of the $(C;\alpha>0)$.

3. An example of $[N;\{p_k\}]$ that includes $(C;\alpha)$ for all $\alpha>0$. If for a fixed $\alpha>0$ the condition

$$(3.1) \qquad c_n(\alpha)\geqq 0 \qquad (n=0,1,2,\cdots),$$

holds for a given regular Nörlund method, it follows at once from (2.3), (2.5), and (2.6) that

$$\sum_{m=0}^{\infty} g_m(\alpha)z^m=(1-z)^{-1}p(z)=\sum_{m=0}^{\infty} P_m z^m.$$

Consequently, $g_m(\alpha)=P_m$ $(m=0,1,2,\cdots)$, so that (2.8) in this event is trivially satisfied. On the other hand, for $n=1$ the relations (2.4) and (3.1) yield $c_1(\alpha)=p_1-\alpha\geqq 0$ or $p_1\geqq\alpha$. Thus if $[N;\{p_k\}]$ is independent of α condition (3.1) can not hold for all values of α. In view of these circumstances we are led to seek some modification of (3.1) and the following one proves to be useful. We shall say that a Nörlund method satisfies *Condition S* if, and only if, (i) it is regular; (ii) independent of α; and (iii) there exists an integer $r=r(\alpha)$ such that the condition

$$(3.2) \qquad c_n(\alpha)\geqq 0 \text{ for all } n\geqq r(\alpha),$$

is satisfied for each $\alpha>0$.

We now state the following theorem.

THEOREM 2. *If $[N;\{p_k\}]$ satisfies Condition S and if*

$$(3.3) \qquad m^\alpha=O(P_m) \qquad (m\to\infty),$$

for each $\alpha>0$, then $[N;\{p_k\}]$ includes $(C;\alpha)$ for all positive α.

Proof. From (3.2), (2.3), and (2.5) it follows that

$$C^*_\alpha(z)=\sum_{n=0}^{r-1} h_n(\alpha)z^n+C_\alpha(z),$$

where $h_n(\alpha)\equiv|c_n(\alpha)|-c_n(\alpha)$ for $(n=0,1,2,\cdots,r-1)$. Then (2.6) becomes

$$\sum_{m=0}^{\infty} g_m(\alpha)z^m=\sum_{m=0}^{\infty}\left\{\sum_{j=0}^{m} h_j(\alpha)\binom{m-j+\alpha}{\alpha}\right\}z^m+P(z),$$

where $P(z)\equiv(1-z)^{-1}p(z)=\Sigma P_m z^m$. Hence for $m\geqq r-1$ we obtain

$$|g_m(\alpha) - P_m| = \left| \sum_{j=0}^{r-1} h_j(\alpha) \binom{m-j+\alpha}{\alpha} \right| \leqq \binom{m+\alpha}{\alpha} \sum_{j=0}^{r-1} |h_j(\alpha)|.$$

Therefore

$$(3.4) \qquad g_m(\alpha) \leqq H_\alpha \binom{m+\alpha}{\alpha} + |P_m| \qquad (m \geqq r-1),$$

where $H_\alpha \equiv \sum_{j=0}^{r-1} |h_j(\alpha)|$. Finally, by means of (3.4) and the well known relation

$$\binom{m+\alpha}{\alpha} \cong \frac{m^\alpha}{\Gamma(\alpha+1)} \qquad (\alpha \neq -1, -2, -3, \cdots; \ m \to \infty),$$

we see that (3.3) implies (2.8). The theorem then follows from the corollary to Theorem 1.

Example. The Nörlund method $[N; \{\cosh k^{\frac{1}{2}}\}]$ *satisfies the conditions of Theorem* 2. For since $p_k \equiv \cosh k^{\frac{1}{2}} > 0$ $(k = 0, 1, 2, \cdots)$ the regularity conditions (1.2) reduce to $p_m = o(P_m)$. But this condition is easily seen to be satisfied since the inequality

$$P_m \geqq 1 + \int_0^m \cosh x^{\frac{1}{2}}\, dx = 2m^{\frac{1}{2}} \sinh m^{\frac{1}{2}} - 2 \cosh m^{\frac{1}{2}} + 3,$$

immediately yields the estimate $p_m/P_m = O(m^{-\frac{1}{2}})$. Part (i) of Condition S is therefore fulfilled.

Since part (ii) is obvious we pass on to part (iii). On account of (2.9) we may clearly restrict α to integral values. We take $r(\alpha) \equiv \alpha$ and set $\phi(x) \equiv \cosh x^{\frac{1}{2}}$ so that $p_k = \phi(k)$ for $(k = 0, 1, 2, \cdots)$. Then in the usual difference notation (2.4) becomes $c_n(\alpha) = \Delta^\alpha \phi(n-\alpha)$ for $n \geqq \alpha$. By a familiar elementary theorem $\Delta^\alpha \phi(n-\alpha) = \phi^{(\alpha)}(n - \alpha\theta)$ where $0 < \theta < 1$. Since the Maclaurin expansion of $\phi(x)$ shows that all of its derivatives are positive for $x > 0$, we conclude that (3.2) is satisfied.

Finally, to see that (3.3) holds we have merely to observe that $P_m^{-1} m^\alpha < p_m^{-1} m^\alpha < 2m^\alpha e^{-m^{\frac{1}{2}}} = o(1)$ for each α as $m \to \infty$.

Consequently, $[N; \{\cosh k^{\frac{1}{2}}\}]$ is an example of a Nörlund method that includes $(C; \alpha)$ for all positive α.

As a final remark it is of interest to notice that Abel summability (A), on the other hand, includes all regular Nörlund methods corresponding to real sequences[6] $\{p_k\}$. Thus we have the relation $(C; \alpha) \subset [N; \{\cosh k^{\frac{1}{2}}\}] \subset (A)$ for all $\alpha > 0$.

MICHIGAN STATE COLLEGE.

---

[6] G. F. Woronoi–J. D. Tamarkin, "Extensions of the notion of the limit of the sum of terms of an infinite series," *Annals of Mathematics*, Second series, vol. 33 (1932), pp. 422-428; in particular, p. 426.

MONOTONE DECOMPOSITIONS OF CONTINUA.*

By M. E. Shanks.

A continuous transformation of a T_1 space (points are closed sets [6, p. 59]) generates a decomposition of the space into disjoint closed sets which fill up the space. Conversely, every such decomposition defines one or more hyperspaces, whose points are the sets of the decomposition, depending on the manner in which these sets are topologized. For each such hyperspace there is a continuous transformation of the given space which takes a point into the set to which it belongs. A discussion of different types of decompositions and their hyperspaces can be found in [6].

Until recently, there had been no investigation of the class of all such decompositions without regard for the different possible topologies that might be put on the hyperspace. In a recent paper B. H. Arnold [1] considered the class of all upper semi-continuous (u. s. c.) decompositions [6, p. 67; 4, p. 123] of a T_1 space and showed that it is possible to construct a space from the decompositions, which are partially ordered, homeomorphic to the given space. In another paper [2] the author obtained results on the semi-linear space of all metrics on a compactum, compatible with its topology, by considering the lattice of all u. s. c. decompositions of the space.

In the same order of ideas, it is natural to consider special types of decompositions and to inquire as to how their structure as an ordered system reflects the topological structure of the space. In this paper there will be considered monotone decompositions of compacta, and especially continua, for the study of whose structure they are particularly well suited.

It is shown that two continua are homeomorphic if and only if their lattices of monotone u. s. c. decompositions are isomorphic under a restricted type of isomorphism. The restriction is essential as it is shown that the lattice of monotone decompositions on a dendrite, or a linear graph, is isomorphic with that for an arc.

A wide class of continua, which we call generalized dendrites, is defined and characterized as those continua for which the monotone u. s. c. decom-

---

* Received February 24, 1944; Revised June 2, 1944; Presented to the American Mathematical Society November 27, 1943. The author wishes to thank G. E. Schweigert and B. H. Arnold for the critical comment and suggestions they have made in conversations and correspondence.

positions form a sublattice of the lattice of all u. s. c. decompositions. This class includes dendrites and Knaster continua, and both types make the lattice of monotones distributive. Finally, a characterization of dendrites is obtained. The paper concludes with some examples and problems.

1. The lattice of monotone decompositions. Denote by $\mathfrak{D}_1(X)$ the class of all decompositions of the compactum X into disjoint closed sets. The sets of a decomposition are called *slices.* Partially order $\mathfrak{D}_1(X)$ as follows: if $D_1, D_2 \in \mathfrak{D}_1(X)$, then $D_1 < D_2$ if each slice of D_2 is *contained in* some slice of D_1. Then, as is easy to show, $\mathfrak{D}_1(X)$ is a complete lattice. Denote by $\mathfrak{D}(X)$ the set of all u. s. c. decompositions of X with the same ordering as before. It is shown in [2] that $\mathfrak{D}(X)$ is also a complete lattice. The join, $D_1 \vee D_2$, of two decompositions is the decomposition whose slices are the intersections of slices of D_1 and D_2. Their meet, $D_1 \wedge D_2$, is the join of all decompositions less than both.

The ordering chosen here is opposite to that used by Arnold and may seem unnatural. Nevertheless, it is natural in the sense that the natural ordering of the quasi-metrics on X induces the above ordering of the decompositions [2].

Denote by $\mathfrak{D}_m(X)$ the subset of $\mathfrak{D}(X)$ consisting of all monotone decompositions, i. e. those whose slices are continua, degenerate or not. Then $\mathfrak{D}_m(X)$ will not in general be a sublattice of $\mathfrak{D}(X)$ since in general the intersection of continua is not a continuum. However $\mathfrak{D}_m(X)$ will be a lattice with respect to the order relation in it. The details are left to the reader.[1]

Since $\mathfrak{D}(X)$ is complete, it has a least element O which has a single slice equal to X, and a greatest element I whose slices are all points. I is always monotone while O is monotone only if X is a continuum. Slices consisting of but a single point are *degenerate.* A decomposition is said to be *simple* if it has but one non-degenerate slice. Let $\mathfrak{D}_{sm}(X)$ denote the class of all simple monotone decompositions, which are necessarily upper semi-continuous.

THEOREM 1. 1. *The continua X and Y are homeomorphic*[2] *if and only*

---

[1] That it is actually a lattice will not be needed in this paper. The join in $\mathfrak{D}_m(X)$ of two monotone decompositions is merely the monotone factor of the monotone-light factorization of their join in $\mathfrak{D}(X)$. In other words, the slice containing any point x is the component containing x of the intersection of the slices of the given decompositions which contain x.

[2] It would be desirable to give lattice conditions on $\mathfrak{D}_m(X)$ which would insure that X be a continuum and then reconstruct the space from the lattice. This is one of a class of unsolved problems of characterizing properties of a space in terms of the decomposition lattice or the monotone decomposition lattice.

if there is an isomorphism of $\mathfrak{D}_m(X)$ *onto* $\mathfrak{D}_m(Y)$ *which carries* $\mathfrak{D}_{sm}(X)$ *onto* $\mathfrak{D}_{sm}(Y)$.

Proof. To each point x of X associate the class ξ of all elements of $\mathfrak{D}_{sm}(X)$ which contain x in their non-degenerate slice. This class, ξ, is maximal with respect to the property that the meet of any finite number of its elements is in $\mathfrak{D}_{sm}(X)$, and in fact in ξ. This characterizes ξ since the slices of any such maximal collection form a maximal collection of continua with the finite intersection property, and hence a common point. Under the isomorphism τ, ξ transforms into a collection η of elements of $\mathfrak{D}_{sm}(Y)$ which is also maximal with respect to the above property, and thus determines a point y of Y which is clearly unique.[3] There is thus defined a transformation T of X into Y. Because τ is an isomorphism it is easy to see that T is one-one and onto Y.

To establish the continuity of T we note that if C is a continuum in X and $C' = T(C)$, then C' is also a continuum. For, C determines a simple monotone decomposition D which belongs to each ξ associated with the points x of C, and similarly for $D' = \tau(D)$ as regards C'. Thus C' is a continuum.

Suppose $x_1, \cdots, x_n, \cdots$, is a sequence of points of X approaching x, and $\lim y_n = y \neq T(x)$, where $y_n = T(x_n)$. If T is not continuous at x this situation prevails for some sequence of points x_n.

Case 1. If X is locally connected at x, or infinitely many x_n are on a locally connected subcontinuum containing x, there is a monotonically decreasing sequence of continua, C_n, closing down on x, such that each C_n contains infinitely many x_n. Thus $C'_n = T(C_n)$ contains infinitely many y_n and hence y, so each C'_n contains y and $T(x)$. But to C_n and C'_n there correspond simple monotone decompositions D_n and D'_n respectively. Since the C_n close down on x the D_n have a join equal to I, while the D'_n will have a join which is a simple monotone decomposition whose non-degenerate slice is the intersection of all the C'_n, and unless $y = T(x)$ this join would not be I.

Case 2. Suppose X is not locally connected at x and no locally connected subcontinuum contains infinitely many x_n. There exist then non-degenerate pairwise disjoint continua, C_n, $(n = 1, 2, \cdots)$, containing[4] x_n and converging to a continuum C containing x, $\lim C_n = C$. Consider $C'_n = T(C_n)$ and suppose that the continua C'_n converge to a continuum[4] C'. If δ is the greatest diameter of C, C_n, $(n = 1, 2, \cdots)$, then as $\delta \to 0$, $T(C)$, C'_n, and C' must also have their diameters approach zero, for otherwise the associated decompositions would not approach I. So, for δ sufficiently small $T(C)$ is distinct from C'. Suppose that this is the case.

---

[3] It must be noted that ξ contains simple monotone decompositions with arbitrarily small slices which are non-degenerate. This insures that the finite intersection property holds. The fact that the slices can be arbitrarily small is reflected in the lattice by the property that ξ contains a sequence increasing, in the sense of the order, to I. Thus η contains such a sequence.

Let D_0 be the monotone decomposition whose non-degenerate slices are C_n, $(n = 1, 2, \cdots)$, and C. Then $D'_0 = \tau(D_0)$ has the non-degenerate slices C'_n, $(n = 1, 2, \cdots)$, C', and $T(C)$. Since $T(C)$ is disjoint from C', one gets another u. s. c. decomposition D_0'' greater than D_0' by allowing all points of $T(C)$ to be point slices. Therefore $\tau^{-1}(D_0'')$ is greater than D_0 and has for its non-degenerate slices C_n, $(n = 1, 2, \cdots)$, but not C. But this decomposition fails to be u. s. c. This completes the proof that the specified isomorphism implies the homeomorphism. The converse is obvious.

An examination of the above proof for the case of locally connected continua gives the following theorem.

THEOREM 1. 2. *The locally connected continua X and Y are homeomorphic if and only if $\mathfrak{D}_{sm}(X)$ is isomorphic to $\mathfrak{D}_{sm}(Y)$.*

If X is not a continuum Theorem 1. 1 is false, for $\mathfrak{D}_m(X)$ reduces to a single element if X is 0-dimensional. Also, the assumption that the simple monotone decompositions correspond cannot be relaxed as is shown by the results of the next section.

2. Factorization theorems.

THEOREM 2. 1. *If $X = X_1 + X_2$, where X, X_1, X_2, are compacta, and X_1 and X_2 have at most a finite[5] set of points in common, then $\mathfrak{D}_m(X)$ is isomorphic to the direct product of $\mathfrak{D}_m(X_1)$ and $\mathfrak{D}_m(X_2)$.*

Proof. If D is any monotone u. s. c. decomposition of X and S any slice of D, then $S = S_1 + S_2$, where $S_i = S \cdot X_i$, $i = 1, 2$. Let D_i, $(i = 1, 2)$, be the monotone decomposition of X_i whose slices are the components of S_i for all S. To see that D_i is u. s. c., suppose that $\liminf S_i^n \cdot S_i \neq 0$, as $n \to \infty$, where S_i^n, S_i are the slices of D_i contained in the slices S_i, S of D respectively. Then lim sup $S_i^n \subset S = S_1 + S_2$. But lim sup $S_i^n \subset X_i$ so that lim sup $S_i^n \subset S_i$.

---

[4] By possibly choosing a subsequence.

[5] The following example shows that $X_1 \cdot X_2$ may not be uncountable. Let X_1 and X_2 be two arcs intersecting in a perfect set non-dense on either arc. It can be shown that the lattice of monotones on X is not modular, while the direct product of the monotones on X_1 and X_2 is distributive. The countable case is open.

We have thus a *natural* mapping, τ, of $\mathcal{D}_m(X)$ into $\mathcal{D}_m(X_1) \times \mathcal{D}_m(X_2)$. To see that τ is a mapping "onto" the direct product, suppose $D_i \in \mathcal{D}_m(X_i)$, $(i = 1, 2)$, and $x_1, \cdots, x_n$ are the points common to X_1 and X_2. If X_1 and X_2 have no points in common the proof is immediate. Each $x_i \in S_i', S_i''$, where S_i', S_i'' are slices of D_1 and D_2 respectively. The set $F = \sum_{n-1} (S_i' + S_i'')$ is the sum of a finite number of continua. Now define a decomposition D of X whose slices are the same as those of D_1 and D_2 together except for S_i', S_i'', $(i = 1, \cdots, n)$. These latter slices are replaced by the components of F. D is obviously monotone and is easily shown to be upper semi-continuous. It is apparent that τ maps D on the product of D_1 and D_2. It is also easy to see that τ is one-one since the intersection of X_1 and X_2 is finite.

COROLLARY. *If A is an arc, then $\mathcal{D}_m(A)$ is isomorphic to $\mathcal{D}_m(A) \times \mathcal{D}_m(A)$.*

THEOREM 2.2. *If A is an arc, then the direct product of countably many $\mathcal{D}_m(A)$ is isomorphic to $\mathcal{D}_m(A)$.*

Proof. For simplicity suppose that A is the unit interval $(0, 1)$. It will suffice to decompose A into a countable set of arcs, $\{A_n\}$, such that each monotone decomposition defines a unique decomposition of each A_n, and conversely. Such a set is the following: $A_1 = (0, 1/2)$, $A_2 = (1/2, 3/4), \cdots, A_n = ((2^n - 1)/2^n, (2^{n+1} - 1)/2^{n+1}), \cdots$.

THEOREM 2.3. *If $\mathcal{D}_m(X)$ is isomorphic to $\mathcal{D}_m(A)$, where A is an arc and X is a continuum, then X is hereditarily locally connected.*[6]

Proof. Let τ be such an isomorphism, $\tau(\mathcal{D}_m(X)) = \mathcal{D}_m(A)$. If X is not hereditarily locally connected it contains a continuum of convergence C_0, which is the limit of disjoint continua C_n, $(n = 1, 2, \cdots)$. Denote by D the monotone u. s. c. decomposition whose only non-degenerate slices are the C_n, and by D_n the simple monotone decomposition whose non-degenerate slice is C_n.

Now the join of D_i and D_j, if $i \neq j$, is I and so also for $\tau(D_i)$ and $\tau(D_j)$. The non-degenerate slices of $\tau(D_i)$ and $\tau(D_j)$ therefore do not overlap. Consider the decomposition, D^*, of A which is the meet of $\tau(D_i)$, $(i = 1, 2, \cdots)$.

---

[6] The converse is false for the continuum of the previous footnote is hereditarily locally connected, and in fact regular.

Since the non-degenerate slices of $\tau(D_i)$, $i \geqq 1$, and $\tau(D_0)$ do not overlap, the same is true for D^* and $\tau(D_0)$. Thus $\tau(D) < D^* < \tau(D_i)$ if $i \geqq 1$. Now D^* is u. s. c. since it is a monotone decomposition of an arc, and so $\tau^{-1}(D^*)$ is also. Thus $D < \tau^{-1}(D^*) < D_i$, if $i \geqq 1$, and the non-degenerate slices of $\tau^{-1}(D^*)$ are the same as those of D_i, $i \geqq 1$, but different if $i = 0$. But then C_0 is not a full slice of $\tau^{-1}(D^*)$ and the latter is not upper semi-continuous. This contradiction completes the proof.

An example exists of a continuum which is hereditarily locally connected, and in fact regular, for which $\mathfrak{D}_m(X)$ is not isomorphic to $\mathfrak{D}_m(A)$.

THEOREM 2. 4. *If X is a linear graph or a dendrite then $\mathfrak{D}_m(X)$ is isomorphic to $\mathfrak{D}_m(A)$, where A is an arc.*

Proof. If X is a linear graph, it is the sum of a finite number of arcs with a finite set of points in common, and one may apply Theorem 2. 1 and its corollary.

If X is a dendrite, it is the sum of a countable set of arcs p_iq_i, $(i = 1, 2, \cdots)$, and a set E of end points, such that p_iq_i is a null sequence, and $p_nq_n \cdot \sum_1^{n-1} p_iq_i = q_n$ [4, p. 89]. Denote by X_n the connected linear graph $\sum_1^n p_iq_i$.

If D_i is a monotone decomposition of p_iq_i, for $(i = 1, 2, \cdots)$, we show that these decompositions determine a decomposition D of X which coincides with D_i on p_iq_i. The decompositions D_i, $(i = 1, \cdots, n)$, determine a unique monotone decomposition D^n of X_n since X_n is a linear graph. If $x \varepsilon X_n$ and is contained in the slice S_n of D^n, the addition of more arcs p_iq_i can only increase the slice containing x. Let $S = \overline{\sum_{i=n,\, n+1, \ldots}^{\infty} S_i}$, and consider all the sets, or "slices" S. These fill up X except possibly for some end points. If an end point is not in some S, then define the slice containing it to be degenerate. The sets S are continua and furthermore are disjoint. For, if S_n, S'_n are disjoint slices of D^n, then S and S' cannot have a point in common. If they were to, this point would not be in X_n and there would exist an arc in X exterior to X_n joining S_n and S'_n, but this would give a simple closed curve in X. Therefore the sets S, regarded as slices, define a monotone decomposition D. It is clear that D is u. s. c., since the slices form a null family. D coincides with D_i on p_iq_i by construction.

Conversely, if D is a monotone decomposition of X, then D defines on p_iq_i a unique decomposition D_i. The slices of D_i which intersect X_n define slices of D^n which are unique. These decompositions D^n determine D as described above. To see this, suppose that the D^n were to determine a decomposition D', different from D. Then D' and D would differ on some arc of X and therefore on one of the arcs $p_i q_i$. This is a contradiction.

3. Generalized dendrites. Up to this point we have not considered the structure of $\mathfrak{D}_m(X)$ in relation to the containing lattice, $\mathfrak{D}(X)$, of all u. s. c. decompositions. We observe that there are two distinct problems; the structure of $\mathfrak{D}_m(X)$ by itself as a partially ordered system, and its structure as a subsystem of $\mathfrak{D}(X)$. In addition, one may embed both in $\mathfrak{D}_1(X)$, the lattice of all decompositions. The reader will easily verify that, while $\mathfrak{D}_m(X)$ as a partially ordered set is a lattice, it is not in general, a sublattice of $\mathfrak{D}(X)$.

Define a continuum X to be a *generalized dendrite* if there is a unique irreducible subcontinuum containing each pair of its points. It is easy to prove that X is a generalized dendrite if and only if the intersection of two continua is either a continuum or is vacuous.

THEOREM 3. 1. *The continuum X is a generalized dendrite if and only if $\mathfrak{D}_m(X)$ is a sublattice of $\mathfrak{D}(X)$.*

Proof. Since the meet of monotone decompositions is monotone, we need only consider joins. Suppose then that X is a generalized dendrite and D_1 and D_2 are monotone u. s. c. decompositions of it. Their join, $D_1 \vee D_2$, is the decomposition whose slices are the intersections of slices of D_1 and D_2. But these intersections are continua.

Conversely, suppose $\mathfrak{D}_m(X)$ is a sublattice of $\mathfrak{D}(X)$. Then the join of two monotone decompositions is monotone, which implies that the intersection of continua is a continuum. Therefore X is a generalized dendrite.

THEOREM 3. 2. *If X is a dendrite, then $\mathfrak{D}_m(X)$ is a distributive sublattice of $\mathfrak{D}(X)$.*

Proof. By a theorem due to S. Bergman, it suffices to show that relative complements are unique [5, p. 75]. This may be done directly, but it is simpler to use Theorems 2. 4 and 3. 1. Then $\mathfrak{D}_m(X)$ is a sublattice of $\mathfrak{D}(X)$, isomorphic to $\mathfrak{D}_m(A)$. It is clear that the structure of $\mathfrak{D}_m(X)$ as a partially ordered system coincides with its structure as a sublattice. To prove its distributivity it is only necessary to show that $\mathfrak{D}_m(A)$ is distributive, or that relative complements are unique. But this latter is not difficult since A is an arc, and is left to the reader.

A continuum, each of whose subcontinua is indecomposable, will be called a *Knaster continuum.* The only example of such a continuum is due to B. Knaster [3].

THEOREM 3. 3. *If X is a Knaster continuum then $\mathfrak{D}_m(X)$ is a distributive sublattice of $\mathfrak{D}(X)$.*

Proof. A lattice is distributive if and only if it does not contain one of the lattices L_1 or L_2 as a sublattice [5, p. 75].

L_1: $D_1 > D_2, D_3, D_4 > D_5$ and
(a) $D_2 \vee D_3 = D_3 \vee D_4 = D_4 \vee D_2 = D_1$,
(b) $D_2 \wedge D_3 = D_3 \wedge D_4 = D_4 \triangle D_2 = D_5$.

L_2: $D_1 > D_2, D_3, D_4 > D_5$, with $D_2 > D_3$ and
(c) $D_2 \vee D_4 = D_3 \vee D_4 = D_1$,
(d) $D_2 \wedge D_4 = D_3 \wedge D_4 = D_5$.

If C_1 and C_2 are subcontinua of a Knaster continuum with a non-empty intersection, then either $C_1 \supset C_2$ or $C_2 \supset C_1$ [3, p. 281]. This implies that the Knaster continuum is a generalized dendrite.

Let S_i, $(i = 1, \cdots, 5)$, be the slices of D_i containing an arbitrary point x. Observe that because of the above property of Knaster continua, to get the meet of two decompositions, one merely chooses the decomposition whose slice containing an arbitrary point x is the greater of the two slices. It is easy to see that this is an u. s. c. decomposition.

Consider L_1. From the ordering of the D_i it follows that $S_1 \subset S_2, S_3, S_4 \subset S_5$, and from (a) that $S_2 \cdot S_3 = S_1$, so either $S_2 \subseteq S_3$ or $S_3 \subseteq S_2$. Suppose the former, then $S_1 = S_2$. But $S_3 \cdot S_4 = S_1 = S_2$, and so $S_4 = S_1 = S_2$. It follows from (b) and the manner in which meets are formed that $S_5 = S_2 + S_3 = S_3 = S_3 + S_4 = S_2 + S_4 = S_2 = S_4$. Thus, $S_i = S_j$, $(i, j = 1, \cdots, 5)$. Since this holds for the slices containing an arbitrary point, $D_i = D_j$, $(i, j = 1, \cdots, 5)$.

Now consider L_2. With the same notation for slices, we have $S_1 \subset S_2, S_3, S_4 \subset S_5$ and $S_2 \subset S_3$. Then from (c), $S_1 = S_2 \cdot S_4 = S_3 \cdot S_4$ and there are three cases to be considered.

(1) $S_2 \subset S_4 \subset S_3$.
(2) $S_2 \subset S_3 \subset S_4$.
(3) $S_4 \subset S_2 \subset S_3$.

In case (1) we get from (d) that $S_4 + S_3 = S_5 = S_3 = S_4 + S_2 = S_4$, and from (c) that $S_2 \cdot S_4 = S_1 = S_2 = S_3 \cdot S_4 = S_4$. Thus $S_2 = S_3$.

In case (2) we get from (c) that $S_2 \cdot S_4 = S_1 = S_2 = S_3 \cdot S_4 = S_3$. Thus $S_2 = S_3$.

In case (3) we get from (d) that $S_3 + S_4 = S_3 = S_5 = S_2 + S_4 = S_2$. Thus $S_2 = S_3$.

So in all cases $S_2 = S_3$ and the two relative complements are equal.

Consider now the system, $\mathfrak{D}_{1m}(X)$, of all monotone decompositions of the continuum X, u. s. c. or not. It is almost immediate that X is hereditarily locally connected if and only if all monotone decompositions are u. s. c., or $\mathfrak{D}_{1m}(X) = \mathfrak{D}_m(X)$. For if there is a continuum of convergence, C, in X, $\lim C_n = C$, then the slices C_n define a monotone decomposition which is not upper semi-continuous. Conversely, if all monotone decompositions are u. s. c., then every sequence of continua is a null sequence and there are no continua of convergence.

THEOREM 3.4. *The continuum X is a dendrite if and only if $\mathfrak{D}_m(X)$ is a sublattice of $\mathfrak{D}(X)$ and one of the following conditions is satisfied:*

(1) $\mathfrak{D}_{1m}(X) = \mathfrak{D}_m(X)$;
(2) *$\mathfrak{D}_m(X)$ is isomorphic to $\mathfrak{D}_m(A)$, where A is an arc;*
(3) *Simple monotone decompositions have complements in $\mathfrak{D}_m(X)$.*

Proof. The sublattice condition makes X a generalized dendrite and (1), (2), and (3) each imply the hereditary local connectivity of X. The only statement to verify is (3). If X were to have a continuum of convergence, C, then it is easy to see that the simple decomposition having C as its non-degenerate slice has no complement.

4. Examples and problems. On the unit interval, $(0,1)$, denote by E the usual Cantor set, obtained by deleting middle thirds. The closures of these deleted sub-intervals define a monotone decomposition, D_E, of $(0,1)$, which will be called the Cantor decomposition. Since D_E has no complement, $\mathfrak{D}_m((0,1))$ is not a Boolean algebra.

It has been shown that $\mathfrak{D}_m(X)$ is distributive for dendrites and Knaster continua, two extreme types. Thus the distributivity of $\mathfrak{D}_m(X)$ need not imply anything about the local connectivity of X. However in the simpler non-locally connected continua $\mathfrak{D}_m(X)$ need not even be modular. Suppose X contains an arc A_0 which is the limit of the distinct arcs A_i, $(i = 1, 2, \cdots)$. For example, X might be the familiar sinusoid. Consider, for simplicity, each arc A_i as the unit interval $(0,1)$, with points on A_i with coördinate a, having as a limit the point on A_0 with coördinate a.

In forming the Cantor decomposition there have been removed after the n-th stage $2^n - 1$ middle thirds. The remaining intervals define a monotone decomposition D_n'. The closures of the removed intervals define a monotone decomposition D_n''. $D_n' \vee D_n'' = I$, and $D_n' \wedge D_n'' = 0$ on A_n.

Let D_4 be a monotone decomposition of X which coincides with D_n' on A_n, $(n = 1, 2, \cdots)$, and has degenerate slices everywhere else. Let D_2 be the monotone decomposition which coincides with D_n'' on A_n, $(n = 1, 2, \cdots)$, and with the Cantor decomposition on A_0. Let D_3 be the monotone decomposition which coincides with D_n'' on A_n, $(n = 1, 2, \cdots)$, and has A_0 as a single slice. All other slices are points. Let D_5 be the monotone decomposition which has exactly A_n, $(n = 0, 1, 2, \cdots)$, as its non-degenerate slices, and let $D_1 = I$. It is easy to see that D_1, D_2, D_3, D_4, and D_5 form the lattice L_2 of the proof of Theorem 3. 3, and hence $\mathfrak{D}_m(X)$ is not modular.

The following problems suggest themselves.

(1) Give a lattice characterization of $\mathfrak{D}_m(A)$.

(2) How does the lattice structure of $\mathfrak{D}_m(X)$, where X is a Knaster continuum, differ from that of $\mathfrak{D}_m(A)$?

(3) Are there continua for which $\mathfrak{D}_m(X)$ is modular but not distributive?

UNIVERSITY OF MISSOURI.

BIBLIOGRAPHY.

1. B. H. Arnold, "Decompositions of a T_1 space," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 768-778.

2. M. E. Shanks, "The space of metrics on a compact metrizable space," *American Journal of Mathematics*, vol. 66 (1944), pp. 461-469.

3. B. Knaster, "Un continu dont tout sous-continu est indecomposable," *Fundamenta Mathematicae*, vol. 3 (1922), pp. 247-286.

4. G. T. Whyburn, "Analytic topology," *Colloquium Publications of the American Mathematical Society*, vol. 28 (1942).

5. G. Birkhoff, "Lattice theory," *Colloquium Publications of the American Mathematical Society*, vol. 25 (1940).

6. P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935).

AXISYMMETRIC HARMONIC VECTORS.*[1]

By MORRIS MARDEN.

1. Introduction. As is well-known, the theory of analytic functions of a complex variable provides an important method for the study of harmonic functions of two real variables. As shown by Bergman[2] the theory of analytic functions may also be used in the study of harmonic functions of three real variables as well as "harmonic vectors" $\boldsymbol{H}(x, y, z)$ i.e. triples $H_1(x, y, z)$, $H_2(x, y, z)$, $H_3(x, y, z)$ of harmonic functions such that the vector

$$\boldsymbol{H}(x, y, z) = H_1(x, y, z)\boldsymbol{i}_x + H_2(x, y, z)\boldsymbol{i}_y + H_3(x, y, z)\boldsymbol{i}_z$$

satisfies the two relations

$$(1.1) \qquad \nabla \cdot \boldsymbol{H} = 0, \qquad \nabla \times \boldsymbol{H} = 0.$$

The essential feature of Bergman's approach is the use of his operator (a generalization of the Whittaker formula):

$$(1.2) \qquad P(f, \Gamma) = P(f, \Gamma, T) = (2\pi i)^{-1} \int_\Gamma f(u, \zeta)\zeta^{-1}d\zeta.$$

Here ζ is a complex variable,

$$(1.3) \qquad u = x + (i/2)[(y + iz)\zeta^{-1} + (y - iz)\zeta],$$

Γ is a closed, rectifiable curve in the ζ-plane and (x, y, z) represents a point in a sufficiently small neighborhood T of some given point (x_0, y_0, z_0). Let S be a region in the u-plane which contains the locus of points u satisfying relation (1.3) when the point (x, y, z) is in T and the point ζ is on Γ. The

---

* Received May 8, 1944.

[1] A large part of the material in this article was obtained while the author was a fellow under the program of Advanced Instruction in Mechanics at Brown University. Part of the material was presented to the American Mathematical Society, September, 1943, under the title of "On Axial Symmetric Flows" and the remainder presented in April, 1944, under the title of "Axisymmetric Harmonic Vectors."

[2] S. Bergman, *Mathematische Zeitschrift*, vol. 24 (1926), pp. 641-669; *Mathematische Annalen*, vol. 99 (1928), pp. 629-658 and vol. 100 (1929), pp. 534-558; *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 163-174. See also E. P. Beckenbach, *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 937-941, in which harmonic vectors are called "Newtonian" vectors.

operator $P(f, \Gamma, T)$ transforms the class of functions $f(u, \zeta)$ which are analytic in u for u in S and continuous in ζ for ζ on Γ, into the class of complex functions $H_1(x, y, z)$ which are harmonic in T. As proved by Bergman, the harmonic functions

$$
\begin{aligned}
(1.4) \qquad H_2(x, y, z) &= P[(i/2)(\zeta + \zeta^{-1})f, \Gamma, T], \\
H_3(x, y, z) &= P[2^{-1}(\zeta - \zeta^{-1})f, \Gamma, T],
\end{aligned}
$$

together with $H_1(x, y, z)$ have the properties that $H_1(x, y, z)\,dx + H_2(x, y, z)\,dy + H_3(x, y, z)\,dz$ is a total differential and that the vector $\boldsymbol{H}(x, y, z) = H_1(x, y, z)\boldsymbol{i}_x + H_2(x, y, z)\boldsymbol{i}_y + H_3(x, y, z)\boldsymbol{i}_z$ satisfies the two relations (1.1).

An important subclass of harmonic functions consists of the axisymmetric harmonic functions $\Phi(x, \rho)$, where $\rho^2 = y^2 + z^2$. Such functions may be generated by the operation $P(f, \Gamma, T)$ applied to the function $f(u, \zeta) = \phi(u)$, where $\phi(u)$ is analytic in S. As an axisymmetric harmonic function, $\Phi(x, \rho)$ may be regarded as the velocity potential of a flow symmetric in the x-axis and thus to it corresponds a so-called Stokes' Stream Function $\Psi(x, \rho)$.[3] As we shall see in the following section,

$$
(1.5) \qquad \Psi(x, \rho)^2 = \rho^2 H_2(x, y, z)^2 + \rho^2 H_3(x, y, z)^2,
$$

where $H_2(x, y, z)$ and $H_3(x, y, z)$ are the functions obtained from formulas (1.4) when $f(u, \zeta) = \phi(u)$.

We are thus led to consider the axisymmetric harmonic vector

$$
(1.6) \qquad \boldsymbol{H}(x, \rho) = \Phi(x, \rho)\boldsymbol{i}_x - \rho^{-1}\Psi(x, \rho)\boldsymbol{i}_\rho
$$

where $\boldsymbol{i}_x$ and $\boldsymbol{i}_\rho$ are the unit vectors along the x and ρ-axes in the meridian plane. In the following sections we shall study the properties of the vector $\boldsymbol{H}(x, \rho)$ in relation to the properties of the analytic function $\phi(u)$ *associated* with $\boldsymbol{H}(x, \rho)$. Physically speaking, our problem is the study of the axisymmetric flows represented by the vector $\boldsymbol{H}(x, \rho)$ in relation to the two-dimensional flows represented by the *associate function* $\phi(u)$.

Thus, in **2**, we shall express $\boldsymbol{H}(x, \rho)$ in terms of $\phi(u)$ and, conversely, $\phi(u)$ in terms of $\boldsymbol{H}(x, \rho)$. In **3**, we shall study the $\boldsymbol{H}(x, \rho)$ corresponding to a $\phi(u)$ having a pole of order n; the corresponding $\boldsymbol{H}(x, \rho)$ has, in general, a branch line along a circle. In **4**, for the purpose of constructing a flow of more practical interest, we shall consider the $\phi(u)$ obtained by superposing certain continuous distributions of poles (multiple sources and multiple vortices). The corresponding $\boldsymbol{H}(x; \rho)$ then represents the flow due to a superposition of multiple source filaments and multiple vortex filaments. We study, in **5**, the $H(x, \rho)$ corresponding to essential singularities in $\phi(u)$ and, in **6**, the $\boldsymbol{H}(x, \rho)$ corresponding to periodic $\phi(u)$. Finally, in **7**, we compute the energy in the field due to any branch of $\boldsymbol{H}(x, \rho)$ in terms of the transforms of the energy in the field with force potential $\phi(u)$.

---

[3] In general, both $\Phi(x, \rho)$ and $\Psi(x, \rho)$ will be complex functions and there is thus an axisymmetric flow whose potential and stream-functions are respectively $\mathfrak{R}(\Phi)$ and $\mathfrak{R}(\Psi)$, and another such flow with these functions as $\mathfrak{J}(\Phi)$ and $\mathfrak{J}(\Psi)$.

2. The vector $\boldsymbol{H}(x, \rho)$ and the associated analytic function $\varphi(u)$. We shall begin by computing $H_1(x, y, z)$, when $f(u, \zeta) = \phi(u)$, where $\phi(u)$ is analytic in a ρ-convex region S; i. e., a region S which, if it contains a point (x, ρ), contains also all of the points $(x, \lambda\rho)$ with $-1 \leqq \lambda \leqq 1$.

$$H_1(x, y, z) = P[\phi(u), |\zeta| = 1] = (2\pi)^{-1} \int_0^{2\pi} \phi(x + iy \cos t + iz \sin t)\, dt$$
$$= (2\pi)^{-1} \int_{0-\alpha}^{2\pi-\alpha} \phi(x + i\rho \cos t)\, dt,$$

where $y = \rho \cos \alpha$ and $z = \rho \sin \alpha$. Hence,

$$(2.1) \qquad H_1(x, y, z) = (2\pi)^{-1} \int_0^{2\pi} \phi(x + i\rho \cos t)\, dt = \Phi(x, \rho).$$

Next, let us evaluate $H_2(x, y, z)\boldsymbol{i}_y + H_3(x, y, z)\boldsymbol{i}_z$.

$$H_2\boldsymbol{i}_y + H_3\boldsymbol{i}_z = P[(i/2)(\zeta + \zeta^{-1})\phi(u), |\zeta| = 1]\boldsymbol{i}_y$$
$$+ P[(1/2)(\zeta - \zeta^{-1})\phi(u), |\zeta| = 1]\boldsymbol{i}_z$$
$$= i(2\pi)^{-1} \int_0^{2\pi} \phi(x + iy \cos t + iz \sin t)(\boldsymbol{i}_y \cos t + \boldsymbol{i}_z \sin t)\, dt.$$

Expanding $\cos(t + \alpha)$ and $\sin(t + \alpha)$, using the fact that $\int_0^{2\pi} \phi(u) \sin t dt = 0$ and substituting $\boldsymbol{i}_\rho = \boldsymbol{i}_y \cos t + \boldsymbol{i}_z \sin t$, we find [4]

$$(2.2) \quad H_2\boldsymbol{i}_y + H_3\boldsymbol{i}_z = i(2\pi)^{-1}\boldsymbol{i}_\rho \int_0^{2\pi} \phi(x + i\rho \cos t) \cos t dt = -\rho^{-1}\Psi(x, \rho)\boldsymbol{i}_\rho$$

and thus

$$\boldsymbol{H}(x, \rho) = \Phi\boldsymbol{i}_x - \rho^{-1}\Psi\boldsymbol{i}_\rho = (2\pi)^{-1} \int_0^{2\pi} \phi(x + i\rho \cos t)(\boldsymbol{i}_x + i\, \boldsymbol{i}_\rho \cos t)\, dt.$$

---

[4] Shortly after our first announcement of this formula (*Bulletin of the American Mathematical Society*, vol. 49 (1943), p. 694, abstract 225), the formula appeared in print in the paper by L. Bers and A. Gelbart in *Quarterly Journal of Applied Mathematics*, vol. 1 (1943), pp. 168-188.

We shall now show that $\Psi(x,\rho)$, as defined by formula (2. 2), is the Stokes' Stream Function corresponding to the velocity potential $\Phi(x,\rho)$ given by formula (2. 1). For this purpose, it is necessary and sufficient to show that $\Phi(x,\rho)$ and $\Psi(x,\rho)$ together satisfy the system of differential equations:

$$(2.\,3)\quad \nabla\cdot\boldsymbol{H}=\rho^{-1}\partial\Psi/\partial\rho-\partial\Phi/\partial x=0,\quad \nabla\times\boldsymbol{H}=(\rho^{-1}\partial\Psi/\partial x+\partial\Phi/\partial x)\boldsymbol{p}=\mathbf{0}$$

where $\boldsymbol{p}$ is a unit vector perpendicular to the meridian plane.

By differentiating equation (2. 2) and then integrating by parts one of the resulting integrals, we find

$$\partial\Psi/\partial\rho=-i(2\pi)^{-1}\int_0^{2\pi}\phi(x+i\rho\cos t)+\rho(2\pi)^{-1}\int_0^{2\pi}\phi'(x+i\rho\cos t)\cos^2 t\,dt$$

$$=\rho(2\pi)^{-1}\int_0^{2\pi}\phi'(x+i\rho\cos t)(\cos^2 t+\sin^2 t)\,dt=\rho(\partial\Phi/\partial x).$$

Similarly, we may verify that $\nabla\times\boldsymbol{H}=\mathbf{0}$.

The above results may be summarized as follows.

THEOREM I. *If $\phi(u)$ is analytic in a ρ-convex region S of the meridian plane, then $\boldsymbol{H}(x,\rho)$, defined by the equations*

$$(2.\,4)\quad \boldsymbol{H}(x,\rho)=(2\pi)^{-1}\int_0^{2\pi}\phi(u)\nabla u\,dt,\ u=x+i\rho\cos t,\ \nabla=\boldsymbol{i}_x\partial/\partial x+\boldsymbol{i}_\rho\partial/\partial\rho,$$

is an axisymmetric harmonic vector in the region obtained by revolving S about the x-axis. The functions $\Phi(x,\rho)$ and $\Psi(x,\rho)$, the "components" of $\boldsymbol{H}(x,\rho)$ according to the formula

$$(2.\,5)\qquad \boldsymbol{H}(x,\rho)=\Phi(x,\rho)\boldsymbol{i}_x-\rho^{-1}\Psi(x,\rho)\boldsymbol{i}_\rho,$$

are, respectively, the velocity potential and Stokes' Stream Function in a three dimensional flow symmetric in the x-axis.[3]

Since $\int_0^{2\pi}\phi(u)\sin t\,dt=0$, a formula alternative to (2. 4) is

$$(2.\,6)\qquad \boldsymbol{H}(x,\rho)=(2\pi)^{-1}\int_0^{2\pi}\phi(u)(\boldsymbol{i}_x+ie^{it}\boldsymbol{i}_\rho)\,dt$$

If, in Theorem I, the region S is taken as the interior and circumference of the circle $r=a$, where $r^2=x^2+\rho^2$, we may write $\phi(u)=\Sigma c_k u^k$ in the circle $r=a$ and then, using the Laplace integrals for the Legendre polynomials $P_n(\cos\theta)$, we find

$$(2.6)\quad \boldsymbol{H}(x,\rho)=(2\pi)^{-1}\int_0^{2\pi}\sum_{k=0}^{\infty}c_k u^k\nabla u\,dt$$
$$=\sum_{k=0}^{\infty}c_k r^k[P_k(x/r)\boldsymbol{i}_x+i(\rho/r)P'_k(x/r)\boldsymbol{i}_\rho].$$

Conversely, in order to determine $\phi(u)$ when $\boldsymbol{H}(x,\rho)$ is given, let us use the orthogonality property of the Legendre functions. Thus,

$$\int_0^{\pi}\boldsymbol{H}(a\cos\theta,a\sin\theta)\cdot[P_n(\cos\theta)\boldsymbol{i}_x-4i\sin\theta P'_n(\cos\theta)\boldsymbol{i}_\rho]\,d\cos\theta$$
$$=2(2n+1)a^n c_n.$$

Hence, if we define the vector

$$\boldsymbol{P}_n(\cos\theta)=(4n+2)^{-1}[P_n(\cos\theta)\boldsymbol{i}_x-4i\sin\theta\,P'_n(\cos\theta)\boldsymbol{i}_\rho]$$

we may write

$$(2.7)\qquad c_n=a^{-n}\int_0^{\pi}\boldsymbol{H}(a\cos\theta,\ a\sin\theta)\cdot\boldsymbol{P}_n(\cos\theta)\,d\cos\theta.$$

If $\boldsymbol{H}(x,\rho)$ is harmonic in $r\leqq a$, it is continuous and therefore its magnitude is bounded in $r\leqq a$. From the properties of Legendre polynomials, it follows that the magnitude of $\boldsymbol{P}_n$ is also bounded. Hence, the $a^n c_n$ are also bounded and the series $\phi(u)=\Sigma c_n u^n$ represents an analytic function at points interior to the circle $r=a$.

We have thus obtained the following result:

THEOREM II. *Let* $\boldsymbol{H}(x,\rho)$ *be an axisymmetric vector which is harmonic in and on the sphere* $x^2+\rho^2=a^2$. *Let* $P_n(\cos\theta)$ *denote the Legendre polynomials of degree n and let*

$$\boldsymbol{P}(u,\theta)=\sum_{n=0}^{\infty}(4n+2)^{-1}(u/a)^n[P_n(\cos\theta)\boldsymbol{i}_x-4i\sin\theta P'_n(\cos\theta)\boldsymbol{i}_\rho].$$

Then an associated analytic function corresponding to $H(x,\rho)$ *for points* (x,ρ) *within the sphere is the function*

$$\phi(x+i\rho)=\int_0^{\pi}\boldsymbol{H}(a\cos\theta,\ a\sin\theta)\cdot\boldsymbol{P}(x+i\rho,\theta)\,d\cos\theta.$$

3. Associate function with a pole. In the following we shall mean by a ρ-convex neighborhood $N_\rho(a,b)$ of a point $u=x+i\rho=a+ib$, $b>0$, the smallest ρ-convex region enclosing the circle $|u-a-ib|=\epsilon$ and by a ρ-convex neighborhood $N'_\rho(a,b)$ that obtained by cutting $N_\rho(a,b)$ along the line segments $x=a$, $b\leqq|\rho|\leqq b+\epsilon$.

If $\phi(x+i\rho)$ is analytic in $N'_\rho(a,b)$ and has a pole of order n at $x+i\rho=a+ib$, then we may write $\phi(u)$ in the form

$$(3.1)\qquad \phi(u)=\sum_1^n c_k(x+i\rho-a-bi)^{-k}+f(x+i\rho)$$

where $f(u)$ is analytic throughout $N_\rho(a,b)$. Formally, the corresponding harmonic vector is

$$(3.2)\qquad \boldsymbol{H}(x,\rho)=\sum_{k=1}^n \boldsymbol{H}_k(x,\rho)+\boldsymbol{F}(x,\rho)$$

where

$$(3.3)\qquad \boldsymbol{H}_k(x,\rho)=(2\pi)^{-1}\int_0^{2\pi}(u-a-ib)^{-k}\nabla u\,dt,\quad u=x+i\rho\cos t,$$

$$\boldsymbol{F}(x,\rho)=(2\pi)^{-1}\int_0^{2\pi}f(u)\nabla u\,dt.$$

Setting $w=e^{it}$ in formula (2.6) and writing

$$2w(x+i\rho\cos t-a-bi)=i\rho(w-w_1)(w-w_2)$$

where

$$(3.4)\qquad w_k=i\rho^{-1}\{x-a-bi+(-1)^k[(x-a-ib)^2+\rho^2]^{1/2}\},$$

we find

$$(3.5)\qquad \boldsymbol{H}_1(x,\rho)=-\pi\rho^{-1}\int_C(w-w_1)^{-1}(w-w_2)^{-1}(\boldsymbol{i}_x+iw\boldsymbol{i}_\rho)\,dw,$$

where C is the unit circle $|w|=1$. Each function $w_k(x,\rho)$ is single-valued in $N'_\rho(a,b)$, and, as $w_1w_2=1$, one of the points w_1, w_2 is inside the unit circle $|w|=1$ and the other is outside this circle, unless both are coincident on this circle. The latter can occur only when the point (x,ρ) is on the radii cut from $N_\rho(a,b)$ in producing $N'_\rho(a,b)$. Let us introduce two sheets over $N'_\rho(a,b)$ and consider the points (x,ρ) of the first sheet to correspond to $|w_1|<1$ and $|w_2|>1$, whereas points (x,ρ) in the second sheet correspond to $|w_1|>1$ and $|w_2|<1$. Then, according to the residue theorem, the value of $\boldsymbol{H}_1(x,\rho)$ on the first sheet is

$$\boldsymbol{H}_1^{(1)}(x,\rho)=2i\rho^{-1}(w_2-w_1)^{-1}(\boldsymbol{i}_x+iw_1\boldsymbol{i}_\rho)$$

and on the second sheet

$$\boldsymbol{H}_1^{(2)}(x,\rho)=2i\rho^{-1}(w_1-w_2)^{-1}(\boldsymbol{i}_x+iw_2\boldsymbol{i}_\rho).$$

From these formulae, since

$$\boldsymbol{H}_n(x,\rho) = (-1)^{n-1}[(n-1)!]^{-1}\,\partial^{n-1}\boldsymbol{H}_1(x,\rho)/\partial x^{n-1},$$

we may obtain the corresponding ones for $\boldsymbol{H}_n(x,\rho)$. In particular, on substituting the values of w_1 and w_2 from (3.4) and setting $x + i\rho - a - ib = \lambda e^{i\theta}$, we learn that for $e^{i\theta} \neq \pm i$ and for $\lambda \to 0$,

$$|\boldsymbol{H}_1(x,\rho)| = O(\lambda^{-\frac{1}{2}}) \text{ or } O(\lambda^{-1}) \text{ according as } b \neq 0 \text{ or } b = 0.$$

Thus

$$|\boldsymbol{H}_n(x,\rho)| = O(\lambda^{\frac{1}{2}-n}) \text{ or } O(\lambda^{-n}) \text{ according as } b \neq 0 \text{ or } b = 0.$$

We may summarize our results as follows.

THEOREM III. *Let $\phi(u)$ be analytic in the neighborhood $N'_\rho(a,b)$ and have a pole of order n at $u = a + ib$. Then the corresponding vector $\boldsymbol{H}(x,\rho)$ is two-valued and harmonic in the region $R[N'_\rho(a,b)]$ obtained on rotating the region $N'_\rho(a,b)$ about the x-axis and has as branch line the circle $x = a$, $\rho = b$. Along this circle, $|\boldsymbol{H}(x,\rho)|$ behaves like $[(x-a)^2 + (\rho-b)^2]^{-\frac{1}{2}(n-\frac{1}{2})}$ when $b > 0$, but like $[(x-a)^2+\rho^2]^{-n/2}$ when $b = 0$.*

4. Circular source and vortex filaments. As is well-known, the principal part of a function $\phi(u)$, having a pole of order n at the point $u = x + i\rho = \xi + i\eta = \zeta$, may be regarded as the complex velocity potential of the two-dimensional flow due to the superposition at $u = \zeta$ of point sources and point vortices of multiplicities from 2 to $n+1$. According to Theorem III, when the associated function $\phi(u)$ has an n-th order pole on the x-axis, the corresponding $\boldsymbol{H}(x,\rho)$ has an n-th order pole at the same point on the x-axis, but, when $\phi(u)$ has a pole at (ξ,η) not on the x-axis, $\boldsymbol{H}(x,\rho)$ has along the circle $x = \xi$, $\rho = \eta$ a line singularity which does not have the character of a multiple circular source or vortex filament. We shall now show that a harmonic vector with a source or vortex filament type of singularity can be derived essentially by two successive applications of the operation (2.4) or (2.6).

More precisely, let us denote the operation (2.4) which transforms a given analytic function $\phi(u,\zeta)$ of $u = x + i\rho$ and $\zeta = \xi + i\eta$ into axisymmetric harmonic vectors thus:

$$(4.1)\quad \boldsymbol{A}_u[\phi(u,\zeta)] = (2\pi)^{-1}\int_0^{2\pi}\phi(u,\zeta)\nabla u\,dt,\ u = x + i\rho\cos t,\ \nabla = \boldsymbol{i}_x\partial/\partial x + \boldsymbol{i}_\rho\partial/\partial\rho$$

$$(4.2)\quad \boldsymbol{A}_\zeta[\phi(u,\zeta)] = (2\pi)^{-1}\int_0^{2\pi}\phi(u,\zeta)\nabla\zeta\,d\alpha,\ \zeta = \xi + i\eta\cos\alpha,\ \nabla = \boldsymbol{i}_x\partial/\partial\xi + \boldsymbol{i}_\rho\partial/\partial\eta.$$

We shall now show that the harmonic vector

$$(4.3) \qquad \boldsymbol{S}_1(x,\rho) = \lambda_1 \boldsymbol{A}_u\{\boldsymbol{i}_x \cdot \boldsymbol{A}_\zeta(u-\zeta)^{-1}\}, \qquad \lambda_1 \text{ real},$$

represents the flow due to a circular source filament of total strength λ_1, along the circle $x = \xi$, $\rho = \eta$, and that the harmonic vector

$$(4.4) \qquad \boldsymbol{V}_1(x,\rho) = \mu_1 \boldsymbol{A}_u\{\boldsymbol{i}_\rho \cdot \boldsymbol{A}_\zeta(u-\zeta)^{-1}\}, \qquad \mu_1 \text{ real},$$

represents the flow due to a circular vortex filament of total strength μ_1 along the same circle $x = \xi$, $\rho = \eta$.

To prove the first statement, let us compute the potential function (Cf. Theorem I)

$$(4.5)\ \boldsymbol{i}_x \cdot \boldsymbol{S}_1(x,\rho) = \lambda_1 \boldsymbol{i}_x \cdot \boldsymbol{A}_u\{\boldsymbol{i}_x \cdot \boldsymbol{A}_\zeta(u-\zeta)^{-1}\}$$

$$= \lambda_1(2\pi)^{-2}\int_0^{2\pi} dt \int_0^{2\pi} (x + i\rho\cos t - \xi - i\eta\cos\alpha)^{-1} d\alpha$$

$$= \lambda_1(2\pi)^{-2}\int_0^{2\pi} dt \int_0^{2\pi} [x + i\rho\cos t - \xi - i\eta\cos(\alpha+t)]^{-1} d\alpha$$

$$= \lambda_1(2\pi)^{-2}\int_0^{2\pi} d\alpha \int_0^{2\pi} [x - \xi + i(\rho - \eta\cos\alpha)\cos t + i\eta\sin\alpha\sin t]^{-1} d$$

$$= \lambda_1(2\pi)^{-1}\int_0^{2\pi} [(x-\xi)^2 + (\rho - \eta\cos\alpha)^2 + \eta^2\sin^2\alpha]^{-\frac{1}{2}} dt.$$

This is precisely the potential due to a circular source filament of total strength λ_1 along the circle $x = \xi$, $\rho = \eta$.

Likewise, let us compute the Stokes' stream function (Cf. Theorem I)

$$(4.6)\ -\rho\boldsymbol{i}_\rho \cdot \boldsymbol{V}_1(x,\rho) = -\rho\mu_1 \boldsymbol{i}_\rho \cdot \boldsymbol{A}_u\{\boldsymbol{i}_\rho \cdot \boldsymbol{A}_\zeta(u-\zeta)^{-1}\}$$

$$= -\rho\mu_1(2\pi)^{-2}\int_0^{2\pi} dt \int_0^{2\pi} [x + i\rho\cos t - i\eta\cos\alpha - \xi]^{-1}\cos(\alpha - t)\, d\alpha$$

$$= -\rho\mu_1(2\pi)^{-2}\int_0^{2\pi} dt \int_0^{2\pi} [x + i\rho\cos t - \xi - i\eta\cos(\alpha+t)]^{-1}\cos\alpha\, d\alpha$$

$$= -\rho\mu_1(2\pi)^{-1}\int_0^{2\pi} [(x-\xi)^2 + (\rho - \eta\cos\alpha)^2 + \eta^2\sin^2\alpha]^{-\frac{1}{2}}\cos\alpha\, d\alpha.$$

This is precisely the Stokes' stream function due to a circular vortex filament of total strength μ_1.

It is to be noted that in the elliptic integrals (4.5) and (4.6) we restrict the points (x,ρ) to the neighborhood $N'_\rho(\xi,\eta)$ and the points (ξ,η) to the neighborhood $N'_\eta(x,\rho)$.

Combining the above results, we see that the harmonic vector

$$(4.7) \qquad \boldsymbol{H}_1(x,\rho) = \boldsymbol{A}_u(\lambda_1\boldsymbol{i}_x + \mu_1\boldsymbol{i}_\rho) \cdot \boldsymbol{A}_\zeta(u-\zeta)^{-1}$$

represents the flow due to the superposition along the circle $x = \xi$, $\rho = \eta$ of a circular source filament of total strength λ_1, and of a circular vortex filament of total strength μ_1.

Let us call a multiple source or vortex point or filament, having a multiplicity of n and having an axis in the direction of the positive x-axis, an *x-oriented n-source* or *n-vortex point* or *filament.* Since the complex velocity potential or harmonic vector for an n-source or n-vortex is the $(n-1)$-st derivative with respect to x of the potential or harmonic vector for a simple source or vortex, we deduce at once, from the preceding discussion, the following theorem:

THEOREM IV. *The axisymmetric harmonic vector $\boldsymbol{H}_n(x, \rho)$ representing the flow due to an x-oriented n-source filament of total strength λ_n superposed along the circle $x = \xi$, $\rho = \eta$ upon an x-oriented n-vortex filament of total strength μ_n is given by the formula:*

$$(4.8) \quad \boldsymbol{H}_n(x, \rho) = (-1)^{n-1}[(n-1)\,!]\boldsymbol{A}_u(\lambda_n \boldsymbol{i}_x + \mu_n \boldsymbol{i}_\rho) \cdot \boldsymbol{A}_\zeta(u-\zeta)^{-n}.$$

It is to be noted that this result is analogous to the well-known expression

$$\phi_n(u) = (-1)^{n-1}(n-1)\,!(\lambda_n + i\mu_n)(u-\zeta)^{-n}.$$

for the complex velocity potential of the two-dimensional flow due to an x-oriented $(n+1)$-source point superposed at $u = \zeta$ upon an x-oriented $(n+1)$-vortex point. Furthermore, in formula (4.8), if $\eta = 0$, we deduce $\boldsymbol{H}_n(x, \rho) = (-1)^{n-1}(n-1)\,!\boldsymbol{A}_u \lambda_n (u-\xi)^{-n}$, the result given at the close of Theorem III.

Since we may write (4.8) as $\boldsymbol{H}_n(x, \rho) = (-1)^{n-1}(n-1)\,!\boldsymbol{A}_u \phi_n(u, \zeta)$ where

$$\begin{aligned} \phi_n(u, \zeta) &= (2\pi)^{-1} \int_0^{2\pi} (\lambda_n + i\mu_n \cos\alpha)(u - \xi - i\eta\cos\alpha)^{-1} d\alpha \\ &= \int_{-\eta}^{+\eta} [\pi^{-1}\lambda_n \eta(\eta^2 - h^2)^{-\frac{1}{2}} + i\pi^{-1}\mu_n(\eta^2 - h^2)^{-\frac{1}{2}}](u - \xi - ih)^{-1} dh, \end{aligned}$$

we may restate Theorem IV in terms of the complex velocity potential due to certain continuous distributions of multiple point sources and vortices along the line segment $x = \xi$, $|\rho| \leqq \eta$.

THEOREM IV′. *Let $\boldsymbol{H}_n(x, \rho)$ be the axisymmetric harmonic vector representing the flow due to an x-oriented n-source filament of total strength λ_n superposed along the circle $x = \xi$, $\rho = \eta$ upon an x-oriented n-vortex filament of total strength μ_n. Then the analytic function $\phi_n(x + i\rho)$ which is the asso-*

ciate of $\boldsymbol{H}_n(x, \rho)$ *is the complex velocity potential due to a continuous distribution of x-oriented* $(n+1)$*-source points of strength* $\pi^{-1}\lambda_n\eta(\eta^2-h^2)^{-\frac{1}{2}}$ *per unit length superposed along the line segment* $x=\xi$, $|\rho| \leqq \eta$ *upon a continuous distribution of x-oriented* $(n+1)$*-vortex points of strength* $\pi^{-1}\mu_n(\eta^2-h^2)^{-\frac{1}{2}}$ *per unit length.*

Finally, from the principal part $\sum_1^n (\lambda_k + i\mu_k)(u-\zeta)^{-k}$ of a function analytic in the neighborhood of $u=\zeta$ except for an n-th order pole at $u=\zeta$, we may derive an analogous "principal part" for an axisymmetric vector which is harmonic in the neighborhood $N'_\rho(\xi, \eta)$ and has a "pole" along the $x=\xi$, $\rho=\eta$; namely,

$$\boldsymbol{A}_u \sum_{k=1}^{n} (\lambda_k \boldsymbol{i}_x + \mu_k \boldsymbol{i}_\rho) \cdot \boldsymbol{A}_\zeta (u-\zeta)^{-k}.$$

The latter is a harmonic vector which represents the flow due to the superposition along the circle $x=\xi$, $\rho=\eta$ of x-oriented k-source and k-vortex filaments of strengths λ_k and μ_k, where $k=1, 2, \cdots, n$.

5. An isolated essential singularity. Let $F(u, \alpha)$ be a function which is continuous in α for $0 \leqq \alpha \leqq 2\pi$ and which for each α is analytic in the neighborhood

$$N(\alpha): \quad 0 < \delta \leqq |u-\xi-i\eta\cos\alpha| \leqq \Delta$$

of the point $u=\xi+i\eta\cos\alpha$, where $F(u, \alpha)$ is assumed to have an isolated essential singularity. Let us suppose also uniform convergence with respect to both u and α for the infinite series portion of the Laurent development of $F(u, \alpha)$:

$$(5.1) \qquad F(u, \alpha) = \sum_{n=1}^{\infty} (a_n + ib_n \cos\alpha)(u-\xi-i\eta\cos\alpha)^{-n} + f(u, \alpha)$$

where $f(u, \alpha)$ is analytic throughout the circle $|u-\xi-i\eta\cos\alpha| \leqq \Delta$. Let us assume further that a_n and b_n are real constants, independent of u and α.

Then

$$\phi(u) = (2\pi)^{-1}\int_0^{2\pi} F(u, \alpha)\, d\alpha = \sum_{n=1}^{\infty} (a_n \boldsymbol{i}_x + b_n \boldsymbol{i}_\rho) \cdot \boldsymbol{A}_\zeta (u-\zeta)^{-n} + f_1(u)$$

is single-valued and analytic in any closed ρ-convex region $N''_\rho(\xi, \eta)$ contained in the neighborhood $N'_\rho(\xi, \eta)$ and the vector

$$(5.2) \qquad \boldsymbol{H}(x, \rho) = \sum_{n=1}^{\infty} \boldsymbol{A}_u (a_n \boldsymbol{i}_x + b_n \boldsymbol{i}_\rho) \cdot \boldsymbol{A}_\zeta (u-\zeta)^{-n} + \boldsymbol{A}_u f_1(u)$$

is single-valued and harmonic in the region $R[N''_\rho(\xi, \eta)]$, obtained by revolving $N''_{\rho}(\xi, \eta)$ about the x-axis. Further, in the torus $0 < \delta \leqq (x-\xi)^2 + (\rho-\eta)^2 \leqq \Delta$, the flow represented by $\boldsymbol{H}(x, \rho)$ will be essentially the flow due to the superposition along the circle $x = \xi$, $\rho = \eta$ of the convergent series of x-oriented n-source and n-vortex filaments represented by the "principal part" of $\boldsymbol{H}(x, \rho)$ in (5.2).

6. Periodic $\boldsymbol{H}(x, \rho)$. For each α, $0 \leqq \alpha \leqq 2\pi$, let $f(u, \alpha)$, $u = x + i\rho$, be a function which is periodic in u with a period of π and which, except for simple poles at $u = \xi_j + i\eta_j \cos \alpha$, $0 \leqq \xi_1 < \xi_2 < \cdots < \xi_n < \pi$, $\eta_j > 0$, is analytic in the strip S: $0 \leqq x \leqq \pi$. Further, let us denote the residue in $u = \xi_j + i\eta_j \cos \alpha$ by $a_j + b_j i \cos \alpha$ and assume that a_j and b_j are independent of α, and that $f(u, \alpha) \rightarrow c \pm di \cos \alpha$ as $\rho \rightarrow \pm \infty$ uniformly with respect to both x and α. Then, as is well known, we may write [5]

$$f(u, \alpha) = c + \sum_{j=1}^{\infty} (a_j + ib_j \cos \alpha) \cot (u - \xi_j - i\eta_j \cos \alpha).$$

In the u-plane, from which has been cut the rays $x = \xi_j + k\pi$, $|\rho| \geqq \eta_j$, $j = 1, 2, \cdots, n$, $k = 0, \pm 1, \pm 2, \cdots$, the function

$$\phi(u) = c + \sum_{j=1}^{\infty} (a_j \boldsymbol{i}_x + b_j \boldsymbol{i}_\rho) \cdot \boldsymbol{A}_\zeta \cot (u - \zeta_j)$$

will be single-valued, analytic and periodic with a period of π and likewise, the harmonic vector $\boldsymbol{H}(x, \rho) = (2\pi)^{-1} \int_0^{2\pi} \phi(u) \nabla u \, dt$ will be single-valued and periodic with a period of π.

The flow represented by this $\boldsymbol{H}(x, \rho)$ will behave, in the neighborhood of each circle $x = \xi_j + k\pi$, $\rho = \eta_j$, $k = 0, \pm 1, \pm 2, \cdots$, like the flow due to the superposition, on the circle, of a circular source filament of total strength a_j and a circular vortex filament of strength b_j. We thus obtain a result similar to the well-known "Kármán Street" of vortices.[6]

7. Energy in an axisymmetric force field.[7] Let

$$\boldsymbol{F}(x, \rho) = X(x, \rho) \boldsymbol{i}_x - {}^{-1}P(x, \rho) \boldsymbol{i}_\rho$$

represent the force in an axisymmetric field and let $f(x + i\rho)$ be the associate function. Then, using a somewhat more general formula than (2.3), we may write

---

[5] Cf. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, American Edition, p. 399, prob. 1.

[6] See L. M. Milne-Thompson, *Theoretical Hydrodynamics*, Macmillan (1938), p. 341.

[7] Cf. S. Bergman, *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 163-174.

$$\mathbf{F}(x,\rho) = (2\pi i)^{-1} \int_{\Gamma} f(u)\,\nabla u \zeta^{-1} d\zeta, \tag{7.1}$$

where

$$u = x + (i/2)\rho(\zeta + \zeta^{-1}), \qquad \nabla = \mathbf{i}_x(\partial/\partial x) + \mathbf{i}_\rho(\partial/\rho) \tag{7.2}$$

and Γ is a rectifiable closed curve in the ζ-plane.

Let us assume in what follows that $f(x + i\rho)$ is an algebraic function having the singularities $x + i\rho = u_1, u_2, \cdots, u_n$ in a ρ-convex region S of the u-plane. Let $\zeta_{jk}(x,\rho)$ be the roots of the equations (7.2) with $u = u_k$ and $k = 1, 2, \cdots, n$. Furthermore, let V be the locus of all points (x,ρ) for which at least two of the $\zeta_{jk}(x,\rho)$ are coincident ($j = 1, 2$; $k = 1, 2, \cdots, n$). The locus V will be called the "singular manifold of $f(u)$." Finally, let the region S' be a ρ-convex region drawn in the region obtained from S by omitting the points common to S and V.

If P_1: (x_1, ρ_1) is chosen as any point of S' and the Γ of formula (7.1) is taken as a curve Γ_1 enclosing one and only point $\zeta_{jk}(x_1,\rho_1)$, say $\zeta_{11}(x_1,\rho_1)$, formula (7.1) will yield a definite branch $\mathbf{F}_1(x,\rho)$ of the many-valued $\mathbf{F}(x,\rho)$. Now beginning at P_1, let us draw in S' a curve C_1 so short that, when (x,ρ) moves from P_1 along C_1, no point $\zeta_{jk}(x,\rho)$ crosses Γ_1. Hence, for all points (x,ρ) on C_1, formula (7.1) with $\Gamma = \Gamma_1$ furnishes the same branch $\mathbf{F}_1(x,\rho)$.

The work W_1, done on a unit particle moving along C_1, in the field $\mathbf{F}_1(x,\rho)$ will then be

$$\begin{aligned} W_1 &= \int_{C_1} \mathbf{F}(x,\rho)\cdot(\mathbf{i}_x dx + \mathbf{i}_\rho d\rho) \\ &= (2\pi i)^{-1} \int_{C_1} (\mathbf{i}_x dx + \mathbf{i}_\rho d\rho) \int_{\Gamma_1} f(u)\,\nabla u \zeta^{-1} d\zeta \\ &= (2\pi i)^{-1} \int_{\Gamma_1} \zeta^{-1} d\zeta \int_{K_1(\zeta)} f(u)\,du, \end{aligned} \tag{7.3}$$

where $K_1(\zeta)$ is the curve described by the point u when ζ is a fixed point on Γ_1 and the point (x,ρ) describes the curve C_1. The quantity

$$w_1(\zeta) = \int_{K_1(\zeta)} f(u)\,du \tag{7.4}$$

represents the work in the two dimensional field $f(u)$. The desired work W_1 is evidently expressed by formula (7.3) as the average of $w_1(\zeta)$ as ζ varies on Γ_1. That is, W_1 is connected with $w_1(\zeta)$ by way of the same operator as given in (7.1); namely, to obtain (7.3), we have only to replace the vector $f(u)\nabla u$ in (7.1) by the scalar $(2\pi i)^{-1}w_1(\zeta)$.

If, instead of the curve C_1, we draw from P_1 in the region S', a curve C of arbitrary length, it will be necessary to establish the following lemma before obtaining results similar to (7.3).

LEMMA. *If C is an arbitrary rectifiable curve drawn from P_1 in S' and if it is divided into a sufficiently large number ν of non-overlapping equal arcs $C_1, C_2, \cdots, C_\nu$, then circles $\Gamma_1, \Gamma_2, \cdots, \Gamma_\nu$ can be drawn in the ζ-plane such that formula (7.1) will furnish the same branch $\boldsymbol{F}_1(x, \rho)$ for all points (x, ρ) on C, provided Γ in (7.1) is taken as Γ_k when the point (x, ρ) is on arc C_k.*

To prove this lemma, let us represent C parametrically in terms of its arc length s; i.e., C: $\omega(s) = x(s)\boldsymbol{i}_x + \rho(s)\boldsymbol{i}_\rho$, $0 \leqq s \leqq l$. Drawn in S', C does not intersect V and thus, for (x, ρ) on C, the points $\zeta_{jk}(s) = \zeta_{jk}[x(s), \rho(s)]$ are distinct. The distance between pairs of points $\zeta_{jk}(s)$ for each s will have a minimum $d > 0$, where d is independent of s. As the $\zeta_{jk}(s)$ are uniformly continuous on C, a $\delta > 0$ exists so that

$$|\zeta_{jk}(s') - \zeta_{jk}(s)| < d/2$$

for all j and k and all s and s' with $|s' - s| \leqq \delta$. Hence, if we let $s_1 = 0$, $s_2 = l/\nu, \cdots, s_\nu = (\nu - 1)l/\nu$, where $\nu > l/\delta$, and if we denote by C_p the arc $s_p \leqq s \leqq s_{p+1}$, then the circle Γ_p, where the equation of Γ_p is

$$|\zeta - \zeta_{11}(s_p)| = d/2, \tag{7.5}$$

will contain no singular point $\zeta_{jk}(s)$ other than $\zeta_{11}(s)$ for $s_p \leqq s \leqq s_{p+1}$. Therefore, if Γ in formula (7.1) is chosen as Γ_p when (x, ρ) is on C_p, formula (7.1) will always provide the same branch $F_1(x, \rho)$.

This lemma enables us to divide C up into a finite number of arcs drawn in S' and to derive for each a result similar to that given by formula (7.3). We may thus state the following theorem.

THEOREM V. *Let C be an arbitrary rectifiable curve in S', P_1: (x_1, ρ_1) any point on C and $\boldsymbol{F}_1(x, \rho)$ the branch of $\boldsymbol{F}(x, \rho)$ furnished by formula (7.1) when Γ encloses one and only singularity $\zeta_{jk}(x, \rho)$.*

Let $K(\zeta)$ be the locus of the point $u = x + (i/2)\rho(\zeta + \zeta^{-1})$ when ζ is fixed and the point (x, ρ) describes C. Let

$$w(\zeta) = \int_{K_1(\zeta)} f(u)\,du \tag{7.6}$$

be the work required to move a unit particle in the two-dimensional field $f(u)$ *along the curve* $K(\zeta)$.

Then ν *circles* $\Gamma_1, \Gamma_2, \cdots, \Gamma_\nu$ *can be drawn in the* ζ*-plane such that the work*

$$W = \int_C \boldsymbol{F}_1(x, \rho) \cdot (\boldsymbol{i}_x dx + \boldsymbol{i}_\rho d\rho) \tag{7.7}$$

required to move a particle in the three-dimensional axisymmetric field $\boldsymbol{F}_1(x, \rho)$ *is the sum* $W = W_1 + W_2 + \cdots + W_n$ *of the transforms of* $w(\zeta)$:

$$W_k = (2\pi i)^{-1} \int_{\Gamma_k} w(\zeta) \zeta^{-1} d\zeta.$$

A simple application of the above theorem is to the case where the force field $\boldsymbol{F}_1(x, \rho)$ is conservative. We may then express $\boldsymbol{F}(x, \rho)$ as the gradient of a potential $\Phi(x, \rho)$ thus: $\boldsymbol{F} = -\nabla\Phi$, $\nabla = \boldsymbol{i}_x\, \partial/\partial x + \boldsymbol{i}_\rho\, \partial/\partial\rho$, where

$$\Phi(x, \rho) = (2\pi i)^{-1} \int_\Gamma \phi(u) \zeta^{-1} d\zeta.$$

Thus, $f(u) = -\phi'(u)$. As the force field $f(u)$ is thus also conservative, we have for the work $w(\zeta)$ corresponding to the arc C_k:

$$w_k(\zeta) = \int_{u_{k+1}}^{u_k} \phi'(u)\, du = \phi(u_k) - \phi(u_{k+1})$$

where $u_k = x_k + (i/2)\rho_k(\zeta + \zeta^{-1})$ and where (x_k, ρ_k) and (x_{k+1}, ρ_{k+1}) are the end-points of C_k. As the Γ_k have been determined so that (7.1) will always furnish the same branch $\boldsymbol{F}_1(x, \rho)$ of the force function, the sum W of the W_k is

$$W = \sum_{k=1}^{\nu} W_k = \Phi(x_1, \rho_1) - \Phi(x_{\nu+1}, \rho_{\nu+1}),$$

a result to be expected in a conservative force field.

UNIVERSITY OF WISCONSIN,
IN
MILWAUKEE.

AN EXTENSION OF A PERRON SYSTEM OF LINEAR EQUATIONS IN INFINITELY MANY UNKNOWNS.*

By I. M. SHEFFER.

---

1. Introduction. Consider the system of equations

$$\sum_{s=0}^{\infty}(a_s+b_{n,n+s})x_{n+s}=c_n, \qquad (n=0,1,\cdots). \tag{1.1}$$

Let the following conditions [(1.2) to (1.5)] be called *conditions* P_1:

$$F(t)\equiv\sum_{s=0}^{\infty}a_st^s \tag{1.2}$$

is analytic in $|t|\leqq 1$;

$$|b_{n,n+s}|\leqq k_n\theta^s, \qquad \theta<1; \tag{1.3}$$

$$k_n\to 0; \tag{1.4}$$

$$a_0+b_{nn}\neq 0, \qquad (n=0,1,\cdots). \tag{1.5}$$

For this system Perron [1] has proved the following theorem:

Let conditions P hold. If $F(t)$ *has* λ *zeros in* $|t|\leqq 1$, *then for every* $\{c_n\}$ *for which* $\limsup|c_n|^{1/n}\leqq 1$ *the general solution* $\{x_n\}$ *of* (1.1) *with* $\limsup|x_n|^{1/n}\leqq 1$ *contains* λ *arbitrary constants. There exist* λ *linearly independent solutions* $\{x_n^{(i)}\}$, $(i=1,\cdots,\lambda)$, *of the homogeneous system with* $\limsup|x_n^{(i)}|^{1/n}\leqq 1$, *and if* $\{x_n^{(0)}\}$ *is a solution of* (1.1) *then the general solution* (*with* $\limsup|x_n|^{1/n}\leqq 1$) *can be written*

$$x_n=x_n^{(0)}+\sum_{i=1}^{\lambda}\alpha_ix_n^{(i)}, \qquad (n=0,1,\cdots), \tag{1.6}$$

where $\alpha_1,\cdots,\alpha_\lambda$ *are arbitrary. Moreover, an integer* M *exists, independent of the sequence* $\{c_n\}$, *with the property that* (1.1) *has one and only one solution* $\{x_n\}$ ($\limsup|x_n|^{1/n}\leqq 1$) *for which* $x_M,x_{M+1},\cdots,x_{M+\lambda-1}$ *are preassigned.*

---

* Received January 3, 1944.

[1] O. Perron, "Über Summengleichungen und Poincarésche Differenzengleichungen," *Mathematische Annalen*, vol. 84 (1921), pp. 1-15.

It is our purpose here to enlarge the Perron system. We shall consider the system of equations

$$(1.7) \qquad \sum_{s=0}^{n-1} b_{ns}x_s + \sum_{s=0}^{\infty} (a_s + b_{n,n+s})x_{n+s} = c_n, \qquad (n = 0, 1, \cdots).$$

As the *limit superior* of a sequence occurs often in the discussion, we introduce the following name and notation.

Definition. By the *type* of a sequence $\{x_n\}$ is meant the number $((x_n))$ given by

$$(1.8) \qquad ((x_n)) \equiv \limsup |x_n|^{1/n}.$$

We impose on the system (1.7) the following conditions [(i) and (ii)] to be referred to as *conditions* Q:

(i) Conditions P hold (save that (1.5) is no longer assumed).

(ii) There exists a number ρ in $0 \leqq \rho < 1$ such that

$$(1.9) \qquad ((\sum_{s=0}^{n-1} b_{ns}x_s)) \leqq \rho$$

for every $\{x_n\}$ for which $((x_n)) \leqq 1$. Condition (ii) is completely equivalent to the following: If

$$(1.10) \qquad C^*_n(t) \equiv \sum_{s=0}^{n-1} |b_{ns}| t^s, \qquad (n = 0, 1, \cdots),$$

then given $\epsilon > 0$, there corresponds a $\delta_\epsilon > 0$ such that

$$(1.11) \qquad ((C^*_n(t))) < \rho + \epsilon$$

for all t in $|t| \leqq 1 + \delta_\epsilon$. (Or, $((C^*_n(1 + \delta_\epsilon))) < \rho + \epsilon$.)

Observe that the system (1.1) is a diagonal system; in other words $x_0, \cdots, x_{n-1}$ occur only in the first n equations. The system (1.7) does not enjoy this property, and therein lies the source of whatever difference there is in the two systems, as we shall see.

The genesis of the Perron system is the linear recurrence relation

$$(1.12) \qquad p_0x_n + p_1x_{n+1} + \cdots + p_\lambda x_{n+\lambda} = c_n, \qquad (n = 0, 1, \cdots),$$

with $p_0p_\lambda \neq 0$. The corresponding homogeneous system has exactly λ linearly independent solutions, and that these are all of finite type is seen by exhibiting a fundamental set of solutions:

$$x_n = t_i^n, nt_i^n, \cdots, n^{k_i-1}t_i^n, \qquad (i = 1, \cdots, \sigma), \tag{1.13}$$

where $t_1, \cdots, t_\sigma$ are the distinct zeros of the polynomial

$$P(t) = p_0 + p_1 t + \cdots + p_\lambda t^\lambda, \tag{1.14}$$

and $k_1, \cdots, k_\sigma$ are the corresponding orders, so that $k_1 + \cdots + k_\sigma = \lambda$. The system (1.12) itself has as a particular solution, corresponding to every sequence $\{c_n\}$ of finite type (say $((c_n)) = R$),

$$x_n = \frac{1}{2\pi i}\int_\Gamma C(1/t)\,\frac{t^{n-1}}{P(t)}\,dt, \tag{1.15}$$

where $C(t) \equiv \sum\limits_0^\infty c_n t^n$ is analytic in $|t| < 1/R$ and where Γ is the circle $|t| = R'$, arbitrary save for the two restrictions:

$$R' \neq |t_i|, \ (i = 1, \cdots, \sigma); \qquad R' > R.$$

From (1.15) we see that $((x_n)) \leqq R'$, and from the arbitrariness of R' $(> R)$ we conclude that (1.12) has a solution of type not exceeding R. Actually, this solution is of type equal to R, for if it were of lesser type, then the left side of (1.12) would be of lesser type; i. e., $((c_n)) < R$, which is a contradiction.

An extension of (1.12), in the homogeneous case, was made by Poincaré,[2] who investigated the asymptotic character of solutions of the system

$$a_{n0}x_n + a_{n1}x_{n+1} + \cdots + a_{n\lambda}x_{n+\lambda} = 0, \qquad (n = 0, 1, \cdots), \tag{1.16}$$

where $a_{ni} \to p_i$, $(i = 0, \cdots, \lambda)$.

The system (1.16) may be regarded as the result of adding a perturbation element to (1.12) (homogeneous case), where the perturbation is small enough to preserve certain properties of the solutions of the original system. Perron introduced his system (1.1), in part, for the purpose of sharpening the conclusions of Poincaré. The system (1.16) is a particular case of (1.1), and we may think of (1.1) as representing the addition of a new perturbation element to (1.12). And in this sense, (1.7) provides still another perturbation.

Concerning the Perron theorem some words are in order as to the *conditions P.* Suppose that $F(t)$ is analytic in $|t| \leqq R$. Then $\sum\limits_{s=0}^\infty a_s x_{n+s}$ converges for every $\{x_n\}$ for which $((x_n)) \leqq R$. The type of $\{x_n\}$ may even be larger, how large depending on how much greater than R is the radius of convergence of (1.2). Since, however, we wish to have a theorem applicable to *every* $F(t)$ analytic in $|t| \leqq R$, we must insist that $((x_n)) \leqq R$. In this case, the sequence $\sum_{s=0}^{\infty} a_s x_{n+s}$ is also of type not exceeding R. Again, to have $y_n \equiv \sum_{s=0}^{\infty} b_{n,n+s} x_{n+s}$ exist for all such $\{x_n\}$, we must assume, *if we insist on a condition of the form* (1.3), (1.4), that $\theta < 1/R$; and then it follows that $((y_n)) \leqq R$. The left side of (1.1) is then a sequence of type not exceeding R, and so we naturally assume the condition $((c_n)) \leqq R$ for $\{c_n\}$.

---

[2] H. Poincaré, "Sur les Equations Linéaires aux Différentielles ordinaires et aux Différences finies," *American Journal of Mathematics*, vol. 7 (1885), pp. 203-258.

If we now make the substitutions

$$c'_n = c'/R^n, \quad x'_n = x_n/R^n, \quad a'_s = R^s a_s, \quad b'_{n,n+s} = R^s b_{n,n+s},$$

then (1.1) assumes the same form with all letters primed; and for this new system the Perron conditions hold. This shows that the conditions as to type are not restrictive. On the other hand, the *nature* of conditions (1.3), (1.4) is a restriction, and it is because of this that the perturbed system has properties analogous to those of the simple system (1.12).

With these preliminary remarks we turn to the system (1.7). We shall apply the method of Perron wherever possible.

2. The $\mathfrak{D}$-system of equations. We shall refer to (1.7) as the $\mathfrak{D}$ system:

$$(2.1) \qquad \mathfrak{D}: \quad \sum_{s=0}^{n-1} b_{ns} x_s + \sum_{s=0}^{\infty} (a_s + b_{n,n+s}) x_{n+s} = c_n, \qquad (n = 0, 1, \cdots);$$

and for $\mathfrak{D}$ we assume conditions Q with the further condition $F(t) \not\equiv 0$.

Let $\epsilon > 0$ be given such that $\rho + \epsilon < 1$. Choose $\delta_\epsilon > 0$ to satisfy (1.11), and δ'_ϵ in $0 < \delta'_\epsilon < \delta_\epsilon$ so that

$$(2.2) \qquad ((C^*_n(t))) < \rho + \epsilon/2$$

for all t in $|t| \leqq |1 + \delta'_\epsilon$. A number R can be found to satisfy the following conditions [(i) to (iii)]:

(i) $R\theta < 1 < R \leqq 1 + \delta'_\epsilon$.

(ii) $F(t)$ and $B_n(t)$ are analytic in $|t| \leqq R$, where

$$(2.3) \qquad B_n(t) = \sum_{s=0}^{\infty} b_{n,n+s} t^{n+s}, \qquad (n = 0, 1, \cdots).$$

(iii) $F(t) \neq 0, \qquad 1 < | t | \leqq R.$

Define $A_n(t)$, $C_n(t)$, $D_n(t)$ by the series

$$A_n(t) = \sum_{s=0}^{\infty} a_s t^{n+s} = t^n F(t), \qquad (n = 0, 1, \cdots), \tag{2.4}$$

$$C_n(t) = \sum_{i=1}^{n-1} b_{n_1 s} t^s, \qquad (n = 0, 1, \cdots), \tag{2.5}$$

$$D_n(t) = A_n(t) + B_n(t) + C_n(t), \qquad (n = 0, 1, \cdots). \tag{2.6}$$

Let Γ_r be the circle $| t | = r$, where r is arbitrary in $1 < r \leqq R$. On Γ_r we have, uniformly,

$$| C_n(t) | \leqq \alpha_r (\rho + \epsilon)^n, \qquad | B_n(t) | \leqq k_n r^n / (1 - r\theta). \tag{2.7}$$

Let $((x_n)) \leqq 1$, so that the series

$$X(t) = \sum_{s=0}^{\infty} x_s t^s \tag{2.8}$$

has a radius of convergence at least as great as one. If r lies in $1 < r < R$, then

$$x_n = \frac{1}{2\pi i} \int_{\Gamma_r} X(1/t) t^{n-1} dt, \qquad (n = 0, 1, \cdots). \tag{2.9}$$

The system (2.1) can then be written

$$c_n = \frac{1}{2\pi i} \int_{\Gamma_r} \frac{D_n(t)}{t} X(1/t) dt = \frac{1}{2\pi i} \int (1/t) X(1/t) \{ C_n(t) + t^n F(t) + B_n(t) \} dt, \qquad (n = 0, 1, \cdots). \tag{2.10}$$

Uniformity of convergence permits term-by-term integration. Conversely, from (2.10) we get (2.1) so that we have

LEMMA 1. *Let conditions Q hold, with* $((c_n)) \leqq 1$. *Every sequence* $\{x_n\}$, *with* $((x_n)) \leqq 1$, *that satisfies* (2.1) *also satisfies* (2.10), *and conversely.*

We now apply a transformation, utilized by Perron, that will replace $F(t)$ by a polynomial.

LEMMA 2. *If* $((u_n)) \leqq a$, *and if* $G(t) = \sum_{s=0}^{\infty} g_s t^s$ *is analytic in* $| t | < g$ *where* $g > a$, *then*

$$v_n = \sum_{s=0}^{\infty} g_s u_{n+s} \tag{2.11}$$

is a sequence of type not exceeding a.

To show this, let σ be chosen in $1/g < \sigma < 1/a$, and let Γ_σ be the circle $|t| = \sigma$. Then

$$v_n = \frac{1}{2\pi i}\int_{\Gamma_\sigma} (1/t)\,G(1/t)\{u_n + u_{n+1}t + \cdots\}dt.$$

Since $|u_n| \leqq A_j \cdot \gamma^n$ for every γ in $a < \gamma < 1/\sigma$, we see that $u_n + u_{n+1}t + \cdots \ll A_j\gamma^n/(1-\gamma t)$ on Γ_σ, so that $((v_n)) \leqq \gamma$; and from the arbitrariness of γ (in the range $\alpha < \gamma < 1/\sigma$) we conclude that $((v_n)) \leqq a$.

Let $F(t)$ have λ zeros in $|t| \leqq 1$, say $t = t_1, \cdots, t_\sigma$, of respective orders $k_1, \cdots, k_\sigma$ so that $k_1 + \cdots + k_\sigma = \lambda$. On defining $P(t)$ by

$$P(t) = \prod_{i=1}^{\sigma} (t - t_i)^{k_i} \equiv p_0 + p_1 t + \cdots + p_\lambda t^\lambda \qquad (p_\lambda = 1) \tag{2.12}$$

we have

$$F(t) = P(t)J(t) \tag{2.13}$$

where $J(t)$ is analytic and different from zero in $|t| \leqq R$. (If $\lambda = 0$ then $P(t) \equiv 1$, $J(t) \equiv F(t)$.)

The function $1/J(t)$ is also analytic in $|t| \leqq R$, since $F(t) \neq 0$ in $1 < |t| \leqq R$, so that if

$$1/J(t) = \sum_{n=0}^{\infty} h_n t^n, \tag{2.14}$$

then $((h_n)) < 1/R$.

In Lemma 2 we may make the choice $G(t) = 1/J(t)$, $\{u_n\} = \{c_n\}$. Then (2.10) becomes

$$\gamma_n = \frac{1}{2\pi i}\int_{\Gamma_r} (1/t)X(1/t)\{t^n P(t) + \sum_{s=0}^{\infty} h_s[B_{n+s}(t) + C_{n+s}(t)]\}dt, \qquad (n = 0, 1, \cdots),$$

where

$$\gamma_n = h_0 c_n + h_1 c_{n+1} + \cdots, \qquad (n = 0, 1, \cdots). \tag{2.15}$$

Absolute and uniform convergence on Γ_r is guaranteed by (2.7); and $((\gamma_n)) \leqq 1$ by Lemma 2.

Define new functions $C_n{}^\dagger(t)$, $B_n{}^\dagger(t)$ by

$$C_n{}^\dagger(t) = \sum_{j=0}^{n-1} \Big(\sum_{s=0}^{\infty} h_s b_{n+s,j}\Big) t^j, \tag{2.16}$$

$$B_n{}^\dagger(t) = \sum_{j=n}^{\infty} \Big(\sum_{s=0}^{\infty} h_s b_{n+s,j}\Big) t^j. \tag{2.17}$$

Each double series converges absolutely and uniformly on Γ_r, and we have the system

$$(2.18)\quad \gamma_n = \frac{1}{2\pi i}\int_{\Gamma_r}(1/t)X(1/t)\{C_n^{\dagger}(t) + t^nP(t) + B_n^{\dagger}(t)\}dt, \qquad (n = 0, 1, \cdots),$$

which, when expanded, becomes the system $\mathscr{D}^{\dagger}$:

$$(2.19)\quad \mathscr{D}^{\dagger}\colon \sum_{s=0}^{n-1}\beta_{ns}x_s + \sum_{s=0}^{\lambda}(p_s + \beta_{n,n+s})x_{n+s} + \sum_{s=\lambda+1}^{\infty}\beta_{n,n+s}x_{n+s} = \gamma_n, \qquad (n = 0, 1, \cdots),$$

where

$$(2.20)\qquad \beta_{nj} = \sum_{s=0}^{\infty}h_sb_{n+s,j}, \qquad (n, j = 0, 1, \cdots).$$

$\mathscr{D}^{\dagger}$ is thus a consequence of $\mathscr{D}$.

The system $\mathscr{D}^{\dagger}$ satisfies a set of conditions Q. To see this, choose ω such that $1 + \delta'_\epsilon < \omega^{-1} \leqq 1 + \delta_\epsilon$ and let Γ_ω be the circle $|t| = \omega$. For $j < n + s$,

$$(2.21)\quad |b_{n+s,j}| = \frac{1}{2\pi i}\int_{\Gamma\omega}C^*_{n+s}(1/t)t^{j-1}dt,$$

$$\sum_{s=0}^{\infty}|h_sb_{n+s,j}| = \frac{1}{2\pi i}\int_{\Gamma\omega}t^{j-1}\{\sum_{s=0}^{\infty}|h_s|C^*_{n+s}(1/t)\}dt \leqq \alpha_\omega A_\epsilon\omega^j(\rho + \epsilon)^n;$$

where $A_\epsilon = \sum_{s=0}^{\infty}|h_s|(\rho + \epsilon)^s$. Hence

$$(2.22)\qquad |\beta_{nj}| \leqq \alpha_\omega A_\epsilon\omega^j(\rho + \epsilon)^n, \qquad j < n;$$

so that for all t in $|t| \leqq 1 + \delta'_\epsilon$ (and therefore for $|t| \leqq R$),

$$C_n^{\dagger *}(t) \equiv \sum_{j=0}^{n-1}|\beta_{nj}|t^j \ll \alpha_\omega A_\epsilon(\rho + \epsilon)^n/(1 - \omega t).$$

Consequently,

$$(2.23)\qquad ((C_n^{\dagger *}(t))) \leqq \rho + \epsilon$$

for all t in $|t| \leqq 1 + \delta'_\epsilon$. In view of the independence of $C_n^{\dagger *}(t)$ relative to ϵ, this is equivalent to

$$(2.24)\qquad ((\sum_{s=0}^{n-1}\beta_{ns}x_s)) \leqq \rho$$

for every $\{x_n\}$ for which $((x_n)) \leqq 1$. That is, the relation (ii) of conditions Q is fulfilled.

Now suppose $j \geqq n$. Then

$$\beta_{nj} = \sum_{s=0}^{j-n}h_sb_{n+s,j} + \sum_{s=j+1}^{\infty}h_{s-n}b_{sj} \equiv I_1 + I_2.$$

we have

$$| I_1 | \leqq \sum_{s=0}^{j-n} | h_s | k_{n+s}\theta^{j-n-s},$$

so that, if θ'' is chosen in $\theta < 1/R < \theta'' < 1$, and if

$$l_n = \max\{k_n, k_{n+1}, \cdots\} \qquad (l_n \to 0),$$

then

$$| I_1 | \leqq B l_n \theta''^{j-n},$$

where $B = \sum_{s=0}^{\infty} | h_s | \theta''^{-s}$. Again,

$$| I_2 | \leqq \sum_{s=j+1}^{\infty} | h_{s-n} b_{sj} |,$$

so that from (2.21), which applies since $j < s$ in b_{sj},

$$| I_2 | \leqq \alpha_\omega \cdot \omega^j A_\epsilon (\rho + \epsilon)^n = \alpha_\omega A_\epsilon [\omega(\rho + \epsilon)]^n \cdot \omega^{j-n}.$$

Since $\omega < 1$, $\omega(\rho + \epsilon) < 1$, it is possible to choose θ' to lie in $\max\{\theta'', \omega\} \leqq \theta' < 1$; and on defining k'_n by

$$k' = B l_n + \alpha_\omega A_\epsilon [\omega(\rho + \epsilon)]^n,$$

then

$$| \beta_{nj} | \leqq k'_n \theta'^{j-n}, \qquad j \geqq n.$$

That is,

$$(2.25) \qquad | \beta_{n,n+s} | \leqq k'_n \theta'^s,$$

with $0 < \theta' < 1$ and $k'_n \to 0$.

In other words, the system $\mathfrak{D}^\dagger$ satisfies a set of conditions Q, with θ replaced by a number θ' having the same property. Any larger number will serve equally well if it is less than one.

The system (2.19) is a consequence of (2.1). But also, starting from (2.19), and using $G(t) = J(t)$, $\{u_n\} = \{\gamma_n\}$ in Lemma 2, it is a simple matter to demonstrate that the system (2.1) results. We accordingly have

LEMMA 3. *Let the system* (2.1) *fulfill conditions* Q, *and let* $((c_n)) \leqq 1$. *Then the system* (2.19) *also has the property* Q (*with* θ' *replacing* θ), *and* $((\gamma_n)) \leqq 1$. *Moreover, every sequence* $\{x_n\}$, *of type not exceeding one, that satisfies* (2.1) *also satisfies* (2.19) *and conversely.*

We now consider a further reduction [Perron, *loc. cit.*].

Definition. Given a system of linear forms (or equations) $\mathfrak{F}: \sum_{j=0}^{\infty} f_{ij} x_j$,

$(j = 0, 1, \cdots)$. By the *truncate of $\mathfrak{F}$ of order m*, denoted by $\mathfrak{F}^{(m)}$, is meant the system obtained from $\mathfrak{F}$ by dropping the first m forms (or equations). In particular, $\mathfrak{F}$ is its own truncate of order zero.

The transformation $\{u_n\}$ to $\{v_n\}$ of Lemma 2 was applied to carry the system $\mathfrak{D}$ into $\mathfrak{D}^\dagger$. Since γ_n does not depend on $c_0, \cdots, c_{n-1}$, we conclude that *the same transformation of $\mathfrak{D}$ to $\mathfrak{D}^\dagger$ carries every truncate of $\mathfrak{D}$ into the truncate of $\mathfrak{D}^\dagger$ of the same order.*

THEOREM 1. *Let the system $\mathfrak{D}$ satisfy conditions Q, with $((c_n)) \leqq 1$. There exists a number ζ in $1 < \zeta$, and an integer $M \geqq 0$, both independent of $\{c_n\}$, such that the M-th truncate system $\mathfrak{D}^{(M)}$ has one and only one solution $\{x_n\}$ of type not exceeding ζ, for which the values $x_0, \cdots, x_{\lambda+M-1}$ are assigned. Moreover, this solution is of type not exceeding one.*

The proof is divided into two parts: uniqueness and existence. (Compare Perron, *loc. cit.*) First we take up uniqueness. Instead of the system $\mathfrak{D}$ we may use the equivalent system $\mathfrak{D}^\dagger$.

Let the non-negative integer M be arbitrary at present. Define $\{\delta_n\}$ by

$$\sum_{s=0}^{\infty} (\delta_s/t^s) \equiv t^\lambda/P(t) = \prod_{i=1}^{\lambda-q} (1 - r_i/t)^{-1}, \tag{2. 26}$$

where we express $P(t)$ in the form

$$P(t) = t^q(t - r_1) \cdots (t - r_{\lambda-q}), \qquad 0 < |\, r_i \,| \leqq 1,$$

and let

$$c'_n = \delta_{n-M}\gamma_M + \delta_{n-M-1}\gamma_{M+1} + \cdots + \delta_0\gamma_n, \qquad n \geqq M. \tag{2. 27}$$

From (2. 26) we obtain the relations

$$\begin{aligned} &\delta_0 p_{\lambda-i} + \delta_1 p_{\lambda-i+1} + \cdots + \delta_i p_\lambda = \begin{cases} 1, & i = 0 \\ 0, & i = 1, \cdots, \lambda; \end{cases} \\ &\delta_n p_0 + \delta_{n+1} p_1 + \cdots + \delta_{n+\lambda} p_\lambda = 0, \qquad (n = 0, 1, \cdots); \end{aligned} \tag{2. 28}$$

so that the system

$$x_{\lambda+n} + \sum_{s=0}^{\lambda-1} \sum_{r=0}^{s} p_{s-r}\delta_{n-M-r}x_{M+s} = c'_n - \sum_{r=M}^{n} \sum_{s=0}^{\infty} \delta_{n-r}\beta_{rs}x_s \qquad (n \geqq M) \tag{2. 29}$$

follows from $\mathfrak{D}^\dagger$ and (2. 27).

The series on the right side converge for all $\{x_n\}$ of type not exceeding R since each equation of (2. 29) is only a linear combination of a finite number of equations of $\mathfrak{D}^\dagger$. Moreover, we can go from (2. 29) back to the M-th order truncate of $\mathfrak{D}^\dagger$ by inversion of these linear combinations in an obvious manner.

The number δ'_ϵ has already been defined. By increasing θ', if necessary, we may assume that it lies in the range $1 < 1/\theta' < 1 + \delta'_\epsilon$. Now choose ζ, ζ_1, ζ_2 to satisfy

$$(2.30) \qquad 1 < \zeta_1 < \zeta < \zeta_2 < 1/\theta' < 1 + \delta'_\epsilon.$$

The following inequality is obvious:

$$\binom{\lambda + n - r - 1}{\lambda - 1} \leqq (\lambda + n)^{\lambda-1} \leqq (\lambda + n)^\lambda, \qquad 0 \leqq r \leqq n.$$

Since λ is fixed, the right hand member is a sequence of type one, so that an integer M_1 can be found such that

$$(2.31) \qquad \frac{\alpha_\omega A_\epsilon}{(1 - \omega\zeta_2)(1 - \rho - \epsilon)} (\lambda + n)^\lambda < \tfrac{1}{4}\zeta_2^{\lambda+n}, \qquad n \geqq M_1.$$

Further, indices M_2, M_3, M_4 exist so that

$$(2.32) \qquad \frac{\alpha_\omega A_\epsilon(\lambda + n)^\lambda}{(1 - \omega\zeta_1)(1 - \rho - \epsilon)} < \tfrac{1}{4}\zeta_1^{\lambda+n}, \qquad n \geqq M_2;$$

$$(2.33) \qquad k'_n < \tfrac{1}{2}(1 - \theta'\zeta_2)(\zeta_2 - 1)^\lambda, \qquad n \geqq M_3;$$

$$(2.34) \qquad k'_n < \tfrac{1}{4}(\zeta_1 - 1)^\lambda(1 - \theta'\zeta_1), \qquad n \geqq M_4.$$

Now choose M as any integer (once chosen, fixed) exceeding $\max\{M_1, M_2, M_3, M_4\}$. Suppose that $\{x_n\}$, $\{y_n\}$ are two sequences of type not exceeding ζ, that each satisfies the truncate system $\mathfrak{D}^{\dagger(M)}$, and that $x_n = y_n$, $0 \leqq n \leqq \lambda + M - 1$. We shall show that $x_n \equiv y_n$. Let $z_n = x_n - y_n$. Then from (2.29) we see that $\{z_n\}$ satisfies the system

$$(2.35) \qquad z_{\lambda+n} = -\sum_{r=M}^{n} \sum_{s=\lambda+M}^{\infty} \delta_{n-r}\beta_{rs}z_s, \qquad n \geqq M.$$

We can write

$$z_{\lambda+n} \equiv Q_1 + Q_2$$

where

$$Q_1 = -\sum_{r=M}^{n} \sum_{s=\lambda+M}^{r-1} \delta_{n-r}\beta_{rs}z_s, \qquad Q_2 = -\sum_{r=M}^{n} \sum_{s=r}^{\infty} \delta_{n-r}\beta_{rs}z_s.$$

Now $((z_n)) \leqq \zeta$, so that a *minimum constant* $C \geqq 0$ exists for which

$$(2.36) \qquad |z_n| \leqq C\zeta_2^n, \qquad n \geqq 0;$$

and since $z_n = 0$ for $n < \lambda + M$, the equality in (2.36) is achieved for some value $n \geqq \lambda + M$. Then

$$| Q_1 | \leqq C \sum_{r=M}^{n} | \delta_{n-r} | C_r^{\dagger *}(\zeta_2) \leqq \frac{C\alpha_\omega A_\epsilon}{1-\omega\zeta_2} \sum_{r=M}^{n} | \delta_{n-r} | (\rho+\epsilon)^r,$$

where $C_r^{\dagger *}$ is defined just before (2.23). From (2.26),

$$\sum_{s=0}^{\infty} (\delta_s/t^s) << \prod_{i=1}^{\lambda} (1-1/t)^{-1} = \sum_{s=0}^{\infty} \binom{\lambda+s-1}{s} (1/t^s),$$

so that [Perron, *loc. cit.*]

$$(2.37) \qquad | \delta_s | \leqq \binom{\lambda+s-1}{s} = \binom{\lambda+s-1}{\lambda-1}.$$

Hence

$$| Q_1 | \leqq \frac{C\alpha_\omega A_\epsilon}{1-\omega f_2} (\lambda+n)^\lambda \cdot \frac{1}{1-\rho-\epsilon} < \tfrac{1}{4} C \zeta_2^{\lambda+n}.$$

Again, from (2.25),

$$\begin{aligned} | Q_2 | &\leqq C \sum_{r=M}^{n} \binom{\lambda+n-r-1}{n-r} k'_r \sum_{s=r}^{\infty} \theta'^{s-r} \zeta_2^s \\ &\leqq \tfrac{1}{2} C(\zeta_2-1)^\lambda \sum_{r=M}^{n} \binom{\lambda+n-r-1}{n-r} \zeta_2^r \leqq \tfrac{1}{2} C(\zeta_2-1)^\lambda \cdot \zeta_2^n \sum_{q=0}^{\infty} \\ &\qquad \binom{\lambda+q-1}{q} \zeta_2^{-q} = (C/2)\zeta_2^{\lambda+n}. \end{aligned}$$

This gives the relation

$$| z_{\lambda+n} | \leqq \tfrac{3}{4} C \zeta_2^{\lambda+n}, \qquad n \geqq M.$$

This is inconsistent with (2.36), since C is minimal, unless $C=0$. Consequently $z_n = 0$ for all n, and the uniqueness is hereby established.

We turn to the problem of existence. Let

$$(2.38) \qquad x_i = X_i, \qquad (i = 0, 1, \cdots, \lambda+M-1)$$

be assigned. Define $x_i^{(0)}$ by

$$(2.39) \qquad x_i^{(0)} = X_i, \quad i < \lambda+M; \qquad x_i^{(0)} = 0, \quad i \geqq \lambda+M.$$

Then define a sequence of approximations $x_i^{(k)}$, $(k = 1, 2, \cdots)$ by

$$(2.40) \qquad x_i^{(k)} = x_i^{(k-1)}, \qquad i < \lambda+M;$$

$$x_{\lambda+n}^{(k)} + \sum_{s=0}^{\lambda-1} \sum_{r=0}^{s} p_{s-r} \delta_{n-M-r} x_{M+s}^{(k-1)} = c'_n - \sum_{r=M}^{n} \sum_{s=0}^{\infty} \delta_{n-r} \beta_{rs} x_s^{(k-1)}, \quad n \geqq M.$$

For $k > 1$ we obtain the inequality

$$(2.41) \qquad | x_{\lambda+n}^{(k)} - x_{\lambda+n}^{(k-1)} | \leqq \sum_{r=M}^{n} \sum_{s=\lambda+M}^{\infty} \binom{\lambda+n-r-1}{n-r} | \beta_{rs} | \; | x_s^{(k-1)} - x_s^{(k-2)} |,$$

and for $k = 1$:

$$(2.42) \quad | x_{\lambda+n}^{(1)} - x_{\lambda+n}^{(0)} | \leqq | c'_n | + \sum_{s=0}^{\lambda-1} \sum_{r=0}^{s} \binom{\lambda + n - M - r - 1}{n - M - r} | p_{s-r} | \, | X_{M+s} |$$
$$+ \sum_{r=M}^{n} \sum_{s=0}^{\lambda+M-1} \binom{\lambda + n - r - 1}{n - r} | \beta_{rs} | \, | X_s |$$
$$\equiv | c'_n | + U_1 + U_2.$$

Now $((\gamma_n)) \leqq 1$. Hence if ξ is chosen in $1 < \xi < \zeta_1$, we have $| \gamma_n | \leqq d\xi^n$ for a suitable d. From (2.27), then,

$$| c'_n | \leqq d\xi^n \sum_{s=0}^{n-M} \binom{\lambda + s - 1}{s} \xi^{-s} \leqq d\xi^n \sum_{s=0}^{\infty} \binom{\lambda + s - 1}{s} \xi^{-s};$$

so that a number Λ_1 exists such that

$$| c'_n | \leqq \Lambda_1 \zeta_1^{n+\lambda}, \qquad n \geqq M.$$

Let

$$\sigma_1 = \max\{| X_0 |, \cdots, | X_{\lambda+M-1} |\}, \qquad \sigma_2 = \max\{| p_0 |, \cdots, | p_\lambda |\}.$$

Then

$$U_1 \leqq \sigma_1 \sigma_2 \sum_{s=0}^{\lambda-1} \sum_{r=0}^{s} \binom{\lambda + n - M - r - 1}{\lambda - 1} \leqq R_1 (\lambda + n)^\lambda \leqq \Lambda_2 \zeta_1^{\lambda+n}$$

for a suitably chosen R_1 and Λ_2. Also,

$$U_2 \leqq \sigma_1 \sum_{r=M}^{n} \binom{\lambda + n - r - 1}{n - r} \{ \sum_{s=0}^{r-1} | \beta_{rs} | + \sum_{s=r}^{\infty} | \beta_{rs} | \}$$
$$\leqq \sigma_1 \sum_{r=M}^{n} \binom{\lambda + n - r - 1}{\lambda - 1} \left\{ \frac{\alpha_\omega A_\epsilon}{1 - \omega} (\rho + \epsilon)^r + k'_r/(1 - \theta') \right\}.$$

The brace is bounded, since it approaches 0 as $r \to \infty$. Hence if σ_3 is an upper bound for it,

$$U_2 \leqq \sigma_1 \sigma_3 \cdot n(\lambda + n)^{\lambda-1} \leqq \sigma_1 \sigma_2 (\lambda + n)^\lambda < \Lambda_3 \zeta_1^{\lambda+n}$$

for a suitable Λ_3.

If, then, $\Lambda = \Lambda_1 + \Lambda_2 + \Lambda_3$ we obtain the inequality

$$(2.43) \qquad | x_{\lambda+n}^{(1)} - x_{\lambda+n}^{(0)} | < \Lambda \zeta_1^{\lambda+n}, \qquad n \geqq M.$$

Using this relation,

$$| x_{\lambda+n}^{(2)} - x_{\lambda+n}^{(1)} | \leqq \Lambda \sum_{r=M}^{n} \binom{\lambda + n - r - 1}{n - r} \{ \sum_{s=\lambda+M}^{r-1} | \beta_{rs} | \zeta_1^s$$
$$+ \sum_{s=r}^{\infty} | \beta_{rs} | \zeta_1^s \} \equiv \Lambda \{ T_1 + T_2 \}.$$

It is readily found that

$$T_1 \leqq \frac{\alpha_\omega A_\epsilon}{1-\omega\zeta} \sum_{r=M}^{n} \binom{\lambda+n-r-1}{\lambda-1} (\rho+\epsilon)^r \leqq \frac{\alpha_\omega A_\epsilon(\lambda+n)^n}{(1-\omega\zeta_1)(1-\rho-\epsilon)} < \tfrac{1}{4}\zeta_1{}^{\lambda+n};$$

and that

$$T_2 \leqq \sum_{r=M}^{n} \binom{\lambda+n-r-1}{n-r} \zeta_1{}^r \cdot k'_r \sum_{q=0}^{\infty} (\theta'\zeta_1)^q \leqq \tfrac{1}{4}(\zeta_1-1)^\lambda \sum_{r=M}^{n} \binom{\lambda+n-r-1}{n-r} \zeta_1{}^r$$

$$\leqq \tfrac{1}{4}(\zeta_1-1)^\lambda \cdot \zeta_1{}^n \sum_{q=0}^{\infty} \binom{\lambda+q-1}{q} \zeta_1{}^{-q} = \tfrac{1}{4}\zeta_1{}^{\lambda+n}.$$

Hence

$$| x_{\lambda+n}{}^{(2)} - x_{\lambda+n}{}^{(1)} | \leqq \tfrac{1}{2}\Lambda\zeta_1{}^{\lambda+n}, \qquad n \geqq M.$$

The argument carries over from one approximation to the next, so that we have in general

$$(2.44) \qquad | x_{\lambda+n}{}^{(k)} - x_{\lambda+n}{}^{(k-1)} | \leqq (1/2^{(k-1)}) \cdot \Lambda\zeta_1{}^{\lambda+n}, \qquad n \geqq M.$$

For fixed n, therefore, $x_{\lambda+n}{}^{(k)}$ has a limit as $k \to \infty$:

$$(2.45) \qquad \lim_{k\to\infty} x_n{}^{(k)} = x_n, \qquad (n = 0, 1, \cdots).$$

(For $n < \lambda + M$, $x_n = X_n$.) This sequence we shall now show is a solution of $\mathfrak{D}^{\dagger(M)}$, or, what is the same thing, of (2.29). For $k < p$,

$$| x_{\lambda+n}{}^{(k)} - x_{\lambda+n}{}^{(p)} | \leqq \Lambda\zeta_1{}^{\lambda+n}\{1/2^k + 1/2^{k+1} + \cdots + 1/2^{p-1}\}.$$

Hence on letting $p \to \infty$:

$$(2.46) \qquad | x_{\lambda+n}{}^{(k)} - x_{\lambda+n} | \leqq (\Lambda/2^{k-1})\zeta_1{}^{\lambda+n}, \qquad n \geqq M.$$

Therefore

$$\left| \sum_{r=M}^{n} \delta_{n-r} \sum_{s=0}^{\infty} \beta_{rs}\{x_s{}^{(k)} - x_s\} \right| \leqq (\Lambda/2^{k-1}) \sum_{r=M}^{n} | \delta_{n-r} | \{ \sum_{s=0}^{\infty} | \beta_r | \zeta_1{}^s \}.$$

Since k enters on the right only in the factor 2^{k-1}, we see that for n fixed,

$$\lim_{k\to\infty} \sum_{r=M}^{n} \delta_{n-r} \sum_{s=0}^{\infty} \beta_{rs} x_s{}^{(k)} = \sum_{r=M}^{n} \delta_{n-r} \sum_{s=0}^{\infty} \beta_{rs} x_s;$$

and on letting $k \to \infty$ in (2.40) we obtain (2.29). That is, $\{x_n\}$ is a solution of (2.29).

There remains the matter of type. Since

$$x_{\lambda+n} = x_{\lambda+n}^{(0)} + \sum_{k=1}^{\infty} \{x_{\lambda+n}^{(k)} - x_{\lambda+n}^{(k-1)}\},$$

therefore

$$|x_{\lambda+n}| \leqq \Lambda\zeta_1^{\lambda+n} \sum_{k=1}^{\infty} 2^{1-k} = 2\Lambda\zeta_1^{\lambda+n},$$

so that $((x_n)) \leqq \zeta_1 < \zeta$. We wish to show that we actually have $((x_n)) \leqq 1$. Suppose $((x_n)) = \mu > 1$. Choose a new ζ, say ζ', in $1 < \zeta' < \mu$. Corresponding to this choice we can then determine an index $M' > M$ so that the truncate system $\mathfrak{D}^{\dagger(M')}$ has a unique solution $\{x'_n\}$, of type not exceeding ζ' for which $x'_0, \cdots, x'_{\lambda+M'-1}$ are preassigned. Choose $x'_i = x_i$, $(i = 0, \cdots, \lambda + M' - 1)$, where $\{x_i\}$ satisfies the original truncate $\mathfrak{D}^{\dagger(M)}$. We then have two sets $\{x_n\}$, $\{x'_n\}$ satisfying $\mathfrak{D}^{\dagger(M')}$, and both of type not exceeding ζ. By uniqueness these must be identical. Hence $((x_n)) \leqq \zeta' < \mu$. As this is a contradiction, the assumption that $((x_n)) > 1$ is untenable. That is, the unique solution of $\mathfrak{D}^{\dagger(M')}$ of type not exceeding ζ is in fact of type not exceeding unity. The proof is now complete.

At this point in the Perron system it is possible to determine $x_{M-1}, \cdots, x_0$, uniquely and in that order, in terms of $x_M, \cdots, x_{\lambda+M-1}$, so that the first M equations of the system are also satisfied. This step-by-step ladder is however not available from $\mathfrak{D}^{\dagger(M)}$ to $\mathfrak{D}^{\dagger}$. In fact, the complete system $\mathfrak{D}^{\dagger}$ (or $\mathfrak{D}$) *may fail to have a solution for some sequence* $\{c_n\}$ *of type not exceeding one.*

This is easy to see if the first M linear forms of $\mathfrak{D}$ are linearly dependent, for in this case the corresponding right hand members, that is, the c_n's, cannot be independent; hence sets $c_0, \cdots, c_{M-1}$ will exist for which the first M equations have no solution. A situation like this is entirely consistent with conditions Q. But even if the first M forms are formally linearly independent, it may yet happen that the conditions imposed on $\{x_n\}$ by the truncate system $\mathfrak{D}^{(M)}$ make the first M forms in effect linearly dependent; and if so, constants $c_0, \cdots, c_{M-1}$ can be found for which there will be no solution. We shall presently give an example illustrating this possibility.

LEMMA 4. *Let the conditions and notation of Theorem* 1 *hold. Let* $x_i = X_i$, $(i = 0, \cdots, \lambda + M - 1)$ *be assigned. Then the unique solution* $\{x_n\}$ *of Theorem* 1 *has the form*

$$x_n = H_n + \sum_{i=0}^{\lambda+M-1} H_{ni} X_i, \qquad n \geqq \lambda + M, \tag{2.47}$$

where all H's are independent of X's, and where H_{ni} *depends only on*

$p_0, \cdots, p_\lambda$ *and* $\{\beta_{ij}\}$, $i \geqq M$, *while* H_n *depends only on these and on* $\{c_i\}$, $i \geqq M$.

To see this, let us return to the successive approximations used in the proof of Theorem 1. We have

$$(2.48) \qquad x_{\lambda+n}^{(1)} - x_{\lambda+n}^{(0)} = x_{\lambda+n}^{(1)} = \nu_{\lambda+n}^{(1)} + \sum_{i=0}^{\lambda+M-1} \nu_{\lambda+n,i}^{(1)} X_i, \qquad n \geqq M,$$

where

$$(2.49) \quad \nu_{\lambda+n}^{(1)} = c'_n; \qquad \nu_{\lambda+n,i}^{(1)} = -\sum_{r=M}^{n} \delta_{n-r}\beta_{ri}, \qquad i < M;$$

$$\nu_{\lambda+n,i}^{(1)} = -\sum_{r=0}^{i-M} p_{i-M-r}\delta_{n-M-r} - \sum_{r=M}^{n} \delta_{n-r}\beta_{ri}, \quad M \leqq i \leqq \lambda + M - 1.$$

And by recurrence, for $k > 1$,

$$(2.50) \qquad x_{\lambda+n}^{(k)} - x_{\lambda+n}^{(k-1)} = \nu_{\lambda+n}^{(k)} + \sum_{i=0}^{\lambda+M-1} \nu_{\lambda+n,i}^{(k)} X_i, \qquad n \geqq M,$$

where

$$(2.51) \qquad \nu_{\lambda+n}^{(k)} = -\sum_{r=M}^{n} \sum_{s=\lambda+M}^{\infty} \delta_{n-r}\beta_{rs}\nu_s^{(k-1)},$$

$$\nu_{\lambda+n,i}^{(k)} = -\sum_{r=M}^{n} \sum_{s=\lambda+M}^{\infty} \delta_{n-r}\beta_{rs}\nu_{si}^{(k-1)}.$$

Hence for $n \geqq M$,

$$(2.52) \quad x_{\lambda+n} = \sum_{k=1}^{\infty} (x_{\lambda+n}^{(k)} - x_{\lambda+n}^{(k-1)}) = \sum_{k=1}^{\infty} \{\nu_{\lambda+n}^{(k)} + \sum_{i=0}^{\lambda+M-1} \nu_{\lambda+n,i}^{(k)} X_i\}.$$

All series involved are absolutely convergent in view of the inequalities obtained in Theorem 1.

COROLLARY. *The coefficients* H_n, H_{ni} *of* (2.47) *are given by*

$$(2.53) \qquad H_n = \sum_{k=1}^{\infty} \nu_n^{(k)}, \qquad H_{ni} = \sum_{k=1}^{\infty} \nu_{ni}^{(k)}, \qquad n \geqq \lambda + M.$$

Consider now a particular system $\mathfrak{D}^\dagger$, fulfilling the conditions of Theorem 1, in which $\beta_{n0} \neq 0$ for all n. Having determined M, and having assigned the values $x_i = X_i$, $i < \lambda + M$, the quantities H_n, H_{ni} are then determined. If the set $\{x_n\}$ is then substituted into the first equation of $\mathfrak{D}^\dagger$ the following condition is obtained:

$$\sum_{i=0}^{\lambda+M-1} X_i\{\beta_{0i} + (\sum_{s=\lambda+M}^{\infty} \beta_{0s}H_{si}) + p_i\} = \gamma_0 - \sum_{s=\lambda+M}^{\infty} \beta_{0s}H_s$$

where $p_i = 0$ for $i > \lambda$. Now the relations

$$H_{s0} = H_{s1} = \cdots = H_{s,\lambda+M-1} = 0 \tag{2.54}$$

cannot hold for all $s \geqq \lambda + M$. For suppose they do hold. From (2.51) we get, on summing from $k = 2$ to ∞:

$$(H_{\lambda+n,i} - \nu_{\lambda+n,i}{}^{(1)}) + \sum_{r=M}^{n} \sum_{s=\lambda+M}^{\infty} \delta_{n-r}\beta_{rs}H_{si} = 0 \qquad (n \geqq M);$$

or, on setting $y_s = H_{si}$, $s \geqq \lambda + M$, and $\nu_{si}{}^{(1)} = r_s$:

$$y_{\lambda+m} + \sum_{r=M}^{n} \sum_{s=\lambda+M}^{\infty} \delta_{n-r}\beta_{rs}y_s = r_{\lambda+n}, \qquad n \geqq M. \tag{2.55}$$

If we further define $y_s = 0$, $s < \lambda + M$, then (2.55) is of type (2.29), with $r_{\lambda+n}$ replacing c'_n. As $((r_{\lambda+n})) \leqq 1$ (a fact easily verified), we conclude that (2.55) has a unique solution of type not exceeding one. Now $\{H_{si}\}$, with i fixed, is such a solution. Hence if (2.54) holds, then $r_{\lambda+n} = 0$, $n \geqq M$. That is,

$$\nu_{\lambda+n,i}{}^{(1)} = 0, \qquad (i = 0, 1, \cdots, \lambda + M - 1;\ n = M, M + 1, \cdots).$$

Therefore from (2.49),

$$\sum_{r=M}^{n} \delta_{n-r}\beta_{ri} = 0, \qquad i < M.$$

On taking $n = M, M + 1, \cdots$ we find, successively, (since $\delta_0 = 1$) that $\beta_{Mi} = 0$, $\beta_{M+1,i} = 0, \cdots$. That is, $\beta_{ni} = 0$ for all $n \geqq M$ and for all i in $0 \leqq i < M$. But our hypothesis is that $\beta_{n0} \neq 0$ for all n. We thus have a contradiction, and conditions (2.54) are untenable.

There exists, therefore, an index $s = s_1 \geqq \lambda + M$, and an i_1 in $0 \leqq i_1 < \lambda + M - 1$, for which $H_{s_1,i_1} \neq 0$. We now revise $\mathcal{D}^\dagger$ by altering its first equation in the following manner: First redefine β_{0,s_1} and β_{0s} ($s \geqq \lambda + M$, $s \neq s_1$) so that

$$\beta_{0,s_1} = 1; \qquad \beta_{0s} = 0, \qquad s \geqq \lambda + M,\ s \neq s_1.$$

The first equation of the new $\mathcal{D}^\dagger$ reduces to

$$\sum_{i=0}^{\lambda+M-1} (\beta_{0i} + \rho_i + H_{s_1,i})X_i = \gamma_0 - H_{s_1}. \tag{2.56}$$

Now redefine β_{0i}, $i < \lambda + M$, so that

$$\beta_{0i} + p_i + H_{s_1,i} = 0, \qquad i < \lambda + M.$$

Since $H_{s_1,i_1} \neq 0$, it follows that for $i = i_1$ the quantity $\beta_{0i} + p_i$ is not zero. This assures us that the first equation in the revised system $\mathfrak{D}^\dagger$ does not have all of its coefficients zero. Now, however, the left side of (2. 56) is zero for all choices of X_i, whereas the right hand side, involving as it does the quantity γ_0, need not be zero. Consequently the system $\mathfrak{D}^\dagger$ does not always have a solution. Moreover, the $M-1$ remaining equations of the first M can be redefined so that there is *formal* linear independence.

To summarize, we have

THEOREM 2. *There exist systems $\mathfrak{D}$ with the following properties:*

(i) *Conditions Q hold.*

(ii) *Every finite set of linear forms of $\mathfrak{D}$ is formally linearly independent.*

(iii) *For at least one sequence $\{c_n\}$ with $((c_n)) \leqq 1$, (in fact, "in general"), there is no solution $\{x_n\}$ of $\mathfrak{D}$ for which $((x_n)) \leqq 1$.*

In view of Theorem 2, it is desirable to know when a system $\mathfrak{D}$ will have a solution $\{x_n\}$ for every $\{c_n\}$. Let the conditions of Theorem 1 hold, so that $\mathfrak{D}^{\dagger(M)}$ has a unique solution for preassigned $x_i = X_i$, $i < \lambda + M$. If we substitute the values x_n as given by (2. 47) into the first M equations of $\mathfrak{D}^\dagger$ there result the relations

$$(2.57)\qquad \sum_{s=0}^{\lambda+M-1} X_s\{\beta_{is} + p_{s-i} + \sum_{r=\lambda+M}^{\infty} \beta_{ir}H_{rs}\} = \gamma_i - \sum_{r=\lambda+M}^{\infty} \beta_{ir}H_r, \qquad i < M,$$

where $p_k = 0$ unless $k = 0, 1, \cdots, \lambda$.

Define A_{is} by

$$(2.58)\qquad A_{is} = \beta_{is} + p_{s-i} + \sum_{r=\lambda+M}^{\infty} \beta_{ir}H_{rs}, \qquad \begin{cases} i = 0, 1, \cdots, M-1; \\ s = 0, 1, \cdots, \lambda + M - 1. \end{cases}$$

Then we can state

THEOREM 3. *Let the system $\mathfrak{D}$ (or $\mathfrak{D}^\dagger$) satisfy conditions Q, and let M be chosen as in Theorem 1. A necessary and sufficient condition that $\mathfrak{D}$ have a solution $\{x_n\}$ with $((x_n)) \leqq 1$ for every sequence $\{c_n\}$ for which $((c_n)) \leqq 1$ is that the matrix*

$$\left\| \begin{matrix} A_{00} & \cdots & A_{0,\lambda+M-1} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ A_{M-1,0} & \cdots & A_{M-1,\lambda+M-1} \end{matrix} \right\|$$

be of rank M; and in this case there exist indices j_n, $(r = 1, \cdots, \lambda)$ in the *range* $0 \leqq j_1 < j_2 < \cdots < j_\lambda \leqq \lambda + M - 1$ *and independent of* $\{c_n\}$, *such that on preassigning the values* $x_{j_r} = X_{j_r}$, $(r = 1, \cdots, \lambda)$, *there is one and only one solution* $\{x_n\}$ *of* $\mathfrak{D}$ *with* $((x_n)) \leqq 1$.

The proof is obvious.

We turn now to the *homogeneous system.* If $c_n \equiv 0$, then $\gamma_n \equiv 0$, so that from (2.27), $c'_n \equiv 0$. This implies that $\nu_{\lambda+n}{}^{(k)} \equiv 0$, $(k = 1, 2, \cdots)$, and therefore $H_n \equiv 0$. Let $x_i = X_i$, $(i = 0, \cdots, \lambda + M - 1)$ be given, and let x_n satisfy the M-th truncate (homogeneous) system. The relation (2.47) reduces to

$$x_n = \sum_{i=0}^{\lambda+M-1} H_{ni} X_i, \qquad n \geqq \lambda + M; \tag{2.59}$$

and (2.57) to

$$\sum_{s=0}^{\lambda+M-1} X_s \{\beta_{is} + p_{s-i} + \sum_{r=\lambda+M}^{\infty} \beta_{ir} H_{rs}\} = 0, \tag{2.60}$$

that is, to

$$\sum_{s=0}^{\lambda+M-1} A_{is} X_s = 0, \qquad (i = 0, 1, \cdots, M - 1). \tag{2.61}$$

The homogeneous system has a solution if and only if (2.61) can be satisfied by suitable choice of $X_0, \cdots, X_{\lambda+M-1}$. We accordingly have

THEOREM 4. *Let the homogeneous system* $\mathfrak{D}$ *(or* $\mathfrak{D}^\dagger$*)* $[c_n \equiv \gamma_n \equiv 0]$ *satisfy conditions* Q, *and let* M *be chosen as in Theorem* 1. *There exist precisely* $\lambda + M - q$ *linearly independent solutions of type not exceeding one, where* q *is the rank of the* A*-matrix of Theorem* 3. *In particular, since* $q \leqq M$, *there exist at least* λ *linearly independent solutions, and if* $\lambda \geqq 1$, *there is at least one solution that is not identically zero.*

THE PENNSYLVANIA STATE COLLEGE.

POWER SERIES WITH BOUNDED COEFFICIENTS.*

By R. J. Duffin and A. C. Schaeffer.

The first part of the present paper concerns entire functions of exponential type; the second part deals with applications of these results to power series which are bounded in a sector of the unit circle.[1]

PART I.

1. We shall say that an entire function $f(z)$ is of exponential type k if there is a constant A such that

$$| f(z) | \leqq A e^{k|z|}.$$

It was shown by Cartwright[2] that if an entire function of exponential type $k < \pi$ is bounded at the integer points $0, \pm 1, \cdots$ then it is bounded on the entire real axis. We shall be concerned with entire functions $f(z)$ which are bounded at a sequence of points $\{\lambda_n\}$. Boas[3] showed that if $| \lambda_n - n | < L < 1/(2\pi^2)$, $(n = 0, \pm 1, \cdots)$ then $f(z)$ is bounded for real z. The condition we impose on the sequence $\{\lambda_n\}$ is that there be two constants Γ and γ such that

$$(1) \qquad | \lambda_n - n | \leqq \Gamma, \qquad (n = 0, \pm 1, \pm 2, \cdots),$$

$$(2) \qquad | \lambda_n - \lambda_m | \geqq \gamma > 0, \qquad n \neq m.$$

---

* Part II received May 2, 1942; Revised and Augmented by Part I, August 31, 1944.

[1] The authors are indebted to the referee, who corrected an error in the original manuscript, and to the editor, who suggested several valuable improvements in the proof.

[2] M. L. Cartwright, "On certain integral functions of order one," *Quarterly Journal of Mathematics* (Oxford Series), vol. 7 (1936), pp. 46-55; A. Pfluger, "On analytic functions bounded at the lattice points," *Proceedings of the London Mathematical Society*, (2), vol. 42 (1936-37), pp. 305-315; A. J. Macintyre, "Laplace's transformation and integral functions," *Proceedings of the London Mathematical Society*, (2), vol. 45 (1938-39), pp. 1-20; R. P. Boas, Jr., "Entire functions bounded on a line," *Duke Mathematical Journal*, vol. 6 (1940), pp. 148-169; R. P. Boas, Jr. and A. C. Schaeffer, "A theorem of Cartwright," *Duke Mathematical Journal*, vol. 9 (1942), pp. 879-883.

[3] R. P. Boas, Jr., *loc. cit.*

2. THEOREM I. *Let* $\{\lambda_n\}$ *be a sequence which satisfies* (1) *and* (2). *If* $f(z)$ *is an entire function which satisfies*

$$|f(z)| \leqq Ae^{k|z|}, \qquad k < \pi, \tag{3}$$

$$|f(\lambda_n)| \leqq 1, \qquad (n = 0, \pm 1, \cdots) \tag{4}$$

then

$$|f(z)| \leqq N(\Gamma, \gamma, k)e^{k|y|}. \tag{5}$$

This theorem is a consequence of a somewhat similar result concerning functions which are regular and of exponential growth in the half-plane $\mathcal{R}(z) \geqq 0$. In this case we suppose that the function is bounded at a sequence of points $\{\lambda_n\}$ which satisfies:

$$|\lambda_n - n| \leqq \Gamma, \qquad n > \Gamma, \tag{6}$$

$$|\lambda_n - \lambda_m| \geqq \gamma > 0, \qquad n \neq m. \tag{7}$$

Here Γ and γ are two positive constants, and we suppose that λ_n is defined only for integers $n > \Gamma$. It then follows that all the points of the sequence $\{\lambda_n\}$ lie in the half-plane $\mathcal{R}(z) > 0$.

THEOREM II. *Let* $\{\lambda_n\}$ *be a sequence which satisfies* (6) *and* (7). *If* $f(z)$ *is regular in the half-plane* $\mathcal{R}(z) \geqq 0$ *and*

$$|f(z)| \leqq e^{k|z|}, \qquad k < \pi, \tag{8}$$

$$|f(\lambda_n)| \leqq 1, \qquad n > \Gamma, \tag{9}$$

then

$$|f(z)| \leqq M(\Gamma, \gamma, k)e^{k|y|}, \qquad x \geqq 0. \tag{10}$$

3. Although in Theorem I the constant A of inequality (3) does not influence the dominant $Ne^{k|y|}$ which is obtained for $f(z)$, the analogous statement is not true in Theorem II. We could not replace inequality (8) by an inequality of the form

$$|f(z)| \leqq Ae^{k|z|}, \qquad k < \pi,$$

without changing the inequality (10). This is shown by the example

$$f(z) = \prod_{\nu=0}^{m} (z - \nu) \cdot \prod_{\nu=1}^{m+1} (z + \nu)^{-1} e^{m-z}.$$

This function is regular and bounded in the right half-plane, and it is bounded by 1 at the integer points, $z = 0, 1, 2, \cdots$. However, $|f(\tfrac{1}{2})|$ $e^{m-\frac{1}{2}}(2m+3)^{-1}(2m+1)^{-1}$, which can be made arbitrarily large by choosing

m large. The following corollary of Theorem II is in some respects a closer analogue of Theorem I than is Theorem II.

COROLLARY I. *Let* $\{\lambda_n\}$ *be a sequence which satisfies* (6) *and* (7). *If* $f(z)$ *is regular in* $\mathfrak{R}(z) \geqq 0$ *and*

$$|f(z)| \leqq Ae^{k|z|}, \qquad k < \pi, \tag{11}$$

$$|f(\lambda_n)| \leqq 1, \qquad n > \Gamma, \tag{12}$$

then

$$|f(x)| \leqq N(\Gamma, \gamma, k), \qquad x \geqq A^2/(\pi - k). \tag{13}$$

We shall also be interested in functions which tend toward zero as z becomes infinite through a suitable sequence of values. The following corollary is a generalization of a result of Young[4] who proved that if

$$f(z) = \int_a^b e^{izt}dg(t), \qquad -\pi < a < b < \pi,$$

where $g(t)$ is of bounded variation and $f(n) \to l$ as $n \to \pm\infty$ then $f(x) \to l$ as $x \to \pm\infty$.

COROLLARY II. *Let* $\{\lambda_n\}$ *be a sequence which satisfies* (6) *and* (7). *If* $f(z)$ *is regular in* $\mathfrak{R}(z) \geqq 0$ *and*

$$|f(z)| \leqq Ae^{k|z|}, \qquad k < \pi, \tag{14}$$

$$f(\lambda_n) = o(1), \qquad n \to \infty, \tag{15}$$

then

$$f(x) = o(1), \qquad x \to \infty. \tag{16}$$

4. Since we shall have several occasions to use the *Phragmén-Lindelöf* principle in various forms we state a lemma which will cover the several cases.

LEMMA I. *If* $f(z)$ *is regular in* $\mathfrak{R}(z) \geqq 0$ *and*

$$|f(z)| \leqq Ae^{k|z|}, \qquad \mathfrak{R}(z) \geqq 0, \tag{17}$$

$$|f(x)| \leqq B, \qquad x \geqq 0, \tag{18}$$

where $B \geqq A$ *then*

$$|f(z)| \leqq Be^{k|y|}, \qquad \mathfrak{R}(z) \geqq 0. \tag{19}$$

If $f(z)$ *is an entire function and*

---

[4] R. C. Young, "The asymptotic behavior of $F(z) = \int_{a-0}^{b+0} e^{zt}dg(t)$," *Mathematische Zeitschrift*, vol. 40 (1936), pp. 292-311.

(20) $$f(z) = O(e^{k|z|})$$

(21) $$|f(x)| \leqq B$$

then

(22) $$|f(z)| \leqq Be^{k|y|}.$$

The second part of this lemma has been proved by Pólya and Szegö[5] using a method which is different from the one employed here. If $f(z)$ satisfies (20) and (21) the function $f(z)e^{ikz}$ is bounded on the positive halves of the real and imaginary axis and is of order one or less in the angular region between them. The Phragmén-Lindelöf principle then shows that it is bounded in the first quadrant. Likewise it is bounded in the second quadrant. Hence

$$|f(z)e^{ikz}| \leqq C$$

in the upper half-plane $y \geqq 0$, where C is some constant. But $f(z)e^{ikz}$ is bounded by B on the real axis so applying the Phragmén-Lindelöf principle to this function with the angular region now being the upper half-plane we find that

$$|f(z)e^{ikz}| \leqq B, \qquad y \geqq 0.$$

Likewise we find that the function $f(z)e^{-ikz}$ is bounded by B in the lower half-plane. This proves (22).

The inequality (19) is proved by applying the Phragmén-Lindelöf principle to the functions $f(z)e^{ikz}$ and $f(z)e^{-ikz}$ in the first and fourth quadrants respectively.

5. In the proof of Theorem II we shall make use of Lemma II, which is weaker than Theorem II, but contains the essential step in the proof of that Theorem.

LEMMA II. *Let* $\{\lambda_n\}$ *be a sequence which satisfies* (6) *and* (7). *If* $f(z)$ *is regular in the half-plane* $\Re(z) \geqq 0$ *and*

(23) $$|f(z)| \leqq e^{k|z|}, \qquad k < \pi,$$

(24) $$|f(\lambda_n)| \leqq 1, \qquad n > \Gamma,$$

(25) $$f(x) = O(1), \qquad x \rightarrow \infty,$$

then

(26) $$|f(x)| \leqq M(\Gamma, \gamma, k), \qquad x \geqq 0.$$

---

[5] Pólya-Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 2, p. 36, problem 202.

The point of this lemma is that if Γ, γ, and k are fixed then the functions which satisfy the conditions of Theorem II may be divided into two classes, namely, (1) those which are unbounded on the positive real axis, and (2) those which are bounded on the positive real axis. The lemma states that all functions of the second class are bounded by a common constant M. We shall show that every function which satisfies the conditions of Theorem II is the limit of a sequence of functions of the second class, and is therefore bounded on the positive real axis. The class (1) is therefore non-existant.

Proof of Lemma II. Let the constants Γ, γ, and k be fixed, and let $f_\nu(z)$, $(\nu = 1, 2, 3, \cdots)$ be any sequence of functions which satisfy the conditions of Lemma II. Let

$$c_\nu = \sup_{0 \leqq x < \infty} |f_\nu(x)|. \tag{27}$$

Now each c_ν is finite according to hypothesis, and it is to be shown that the c_ν are bounded in their set. We suppose the sequence $f_\nu(z)$ chosen such that $c_\nu > \nu$, and show that this leads to a contradiction. Each $f_\nu(z)$ is bounded by 1 at a sequence $\{\lambda_n^{(\nu)}\}$ which satisfies (6) and (7). If we combine (23) and (27) with Lemma I we find that

$$|f_\nu(z)| \leqq c_\nu e^{k|y|}.$$

There will be some point at which $|f_\nu(x)|$ is near its upper bound, let x_ν be chosen such that

$$|f_\nu(x_\nu)| \geqq c_\nu(1 - 1/\nu).$$

In view of (23) it follows that $x_\nu \to \infty$ as ν does.

Let

$$\phi_\nu(z) = f_\nu(z + [x_\nu])/c_\nu$$

where $[x_\nu]$ is the integer part of x_ν. Then $\phi_\nu(z)$ is regular in the half plane $\mathcal{R}(z) \geqq -[x_\nu]$ and it satisfies

$$|\phi_\nu(z)| \leqq e^{k|y|}, \qquad x \geqq -[x_\nu],$$

$$\max_{0 \leqq x \leqq 1} |\phi_\nu(x)| \geqq 1 - 1/\nu$$

$$|\phi_\nu(\mu_n^{(\nu)})| \leqq 1/c_\nu, \qquad n > \Gamma - [x_\nu].$$

Here $\{\mu_n^{(\nu)}\}$ is a sequence which is obtained by translating the sequence $\{\lambda_n^{(\nu)}\}$ a distance $[x_\nu]$ to the left.

6. Since the functions $\phi_\nu(z)$ are dominated by $e^{k|y|}$ they are equicontinuous in any given bounded region for large ν, and so Ascoli's convergence theorem concerning bounded equicontinuous functions shows that there is a subsequence which tends to a limit $\phi(z)$ which is an entire function and satisfies

$$(28) \qquad |\phi(z)| \leqq e^{k|y|}; \qquad \max_{0\leqq x\leqq 1} |\phi(x)| = 1.$$

Moreover, the sequence may be so chosen that, for each n, $\mu_n{}^{(\nu)}$ tends to a limit as ν becomes infinite. Then

$$(29) \qquad \phi(\mu_n) = 0, \qquad (n = 0, \pm 1, \cdots)$$

and the sequence $\{\mu_n\}$ satisfies

$$\begin{aligned} &|\mu_n - n| \leqq \Gamma, && (n = 0, \pm 1, \pm 2, \cdots) \\ &|\mu_n - \mu_m| \geqq \gamma > 0, \quad n \neq m. && \end{aligned}$$

But Jensen's formula shows that if $\phi(z)$ is a function which satisfies (28) then

$$\overline{\lim_{r\to\infty}}\, r^{-1} \int_1^r n(t)t^{-1}dt \leqq 2k/\pi < 2$$

where $n(t)$ is the number of zeros of $\phi(z)$ in $|z| \leqq t$. But from (29) it is easily seen that

$$\underline{\lim_{r\to\infty}}\, r^{-1} \int_1^r n(t)t^{-1}dt \geqq 2.$$

This contradiction proves that there exists a constant $M(\Gamma, \gamma, k)$ such that the inequality (26) is valid.

7. Proof of Theorem II. Let $f(z)$ satisfy the conditions of Theorem II and let

$$\rho = \overline{\lim_{x\to\infty}}\, x^{-1} \log |f(x)|.$$

We choose a constant a subject to the condition that $a > (\rho + |\rho|)/2$. Then let $\epsilon > 0$ be chosen such that $\epsilon < a$ and $\epsilon < (\pi - k)/2$. For $(\nu = 1, 2, 3, \cdots)$ let

$$g_\nu(z) = f(z)e^{-az} \sum_{p=0}^{\nu} (\epsilon z)^p/p!.$$

The sum is no greater in modulus than $e^{\epsilon|z|}$ since it is dominated term by term by the Taylor's expansion of $e^{\epsilon|z|}$ about the origin. We then see that in the half-plane $\mathcal{R}(z) \geqq 0$,

$$|g_\nu(z)| \leqq |f(z)|\, e^{\epsilon|z|} \leqq e^{(k+\epsilon)|z|} \leqq e^{\kappa|z|}$$

where we define κ as $(k + \pi)/2 < \pi$. Also if we write $\lambda_n = \mu_n + i\tau_n$ then

$$|g_\nu(\lambda_n)| \leqq e^{-a\mu_n}e^{\epsilon|\lambda_n|} \leqq e^{\epsilon\Gamma}$$

since $\epsilon|\lambda_n| - a\mu_n \leqq \epsilon(\mu_n + |\tau_n|) - \epsilon\mu_n = \epsilon|\tau_n| \leqq \epsilon\Gamma$.

Moreover, $g_\nu(x)$ tends to zero as x becomes infinite with a, ϵ, and ν fixed, and, in particular, $g_\nu(x)$ is bounded on the positive real axis. We see from the above remarks that the function $e^{-\epsilon\Gamma}g_\nu(z)$ satisfies the conditions of Lemma II with k replaced by κ. It therefore follows that

$$(30) \qquad |g_\nu(x)| \leqq e^{\epsilon\Gamma}M(\Gamma, \gamma, \kappa), \qquad x \geqq 0.$$

Now keeping a and ϵ fixed and letting ν become infinite we see that the inequality (30) must also be satisfied by the limit $f(z)e^{-az}e^{\epsilon z}$ of $g_\nu(z)$. Thus

$$(31) \qquad |f(x)| \leqq e^{(a-\epsilon)x}e^{\epsilon\Gamma}M(\Gamma, \gamma, \kappa), \qquad x \geqq 0.$$

We now distinguish two cases, $\rho > 0$ and $\rho \leqq 0$. If $\rho > 0$ we choose an ϵ, $0 < \epsilon < \rho$, of course keeping the previous condition $\epsilon < (\pi - k)/2$. Now keeping ϵ fixed and letting a approach ρ we see that

$$|f(x)| \leqq e^{(\rho-\epsilon)x}e^{\epsilon\Gamma}M(\Gamma, \gamma, \kappa)$$

on the positive real axis. But this contradicts the definition of ρ and shows that $\rho \leqq 0$.

Since $\rho \leqq 0$ we let a and ϵ each approach zero. Then the inequality (31) shows that $f(x)$ is bounded by $M(\Gamma, \gamma, \kappa)$. Since $f(x)$ is bounded by some constant we may now apply Lemma II, and it follows that $f(x)$ is bounded by $M(\Gamma, \gamma, k)$ for $x \geqq 0$. Now if we apply Lemma I to $f(z)$ we see, since clearly $M \geqq 1$, that the inequality (10) is valid throughout the right half-plane.

8. To prove Corollary I it is sufficient to consider the case in which the constant A of inequality (11) is greater than 1, since if $A \leqq 1$ the result is clear from Theorem II. Let $\beta = 2(\log A)/(\pi - k)$ and $\tau = \beta A$. The function

$$g(z) = z(z + \tau)^{-1}f(z)$$

satisfies in the right half-plane $\mathfrak{R}(z) \geqq 0$ the inequalities

$$|g(z)| \leqq \beta\tau^{-1}|f(z)| \leqq A^{-1}Ae^{k|z|} = e^{k|z|}, \qquad |z| \leqq \beta,$$

and

$$|g(z)| \leqq |f(z)| \leqq Ae^{k|z|} \leqq e^{\kappa|z|}, \qquad |z| \geqq \beta,$$

where $\kappa = (\pi + k)/2 < \pi$. Here we have used the fact that $|z(z + \tau)^{-1}| \leqq \beta\tau^{-1} = A^{-1}$ in the semicircle $|z| \leqq \beta$, $\mathfrak{R}(z) \geqq 0$, and $|z(z + \tau)^{-1}| \leqq 1$ in the half-plane $\mathfrak{R}(z) \geqq 0$.

It is thus seen that $g(z)$ satisfies the conditions of Theorem II with k replaced by κ, and so

$$| g(x) | \leqq M(\Gamma, \gamma, \kappa), \qquad x \geqq 0.$$

Then expressing $f(z)$ is terms of $g(z)$ and using the last inequality we see that

$$| f(x) | \leqq (x + \tau) x^{-1} M(\Gamma, \gamma, \kappa) \leqq 2M(\Gamma, \gamma, \kappa)$$

when $x \geqq \tau = \beta A = 2A(\log A)/(\pi - k)$. Since $2 \log A < A$ the result follows if we write $N(\Gamma, \gamma, k)$ in place of $2M(\Gamma, \gamma, \pi/2 + k/2)$.

9. Proof of Corollary II. If $f(z)$ satisfies the conditions of Corollary II let

$$| f(\lambda_n) | < \epsilon, \qquad n \geqq m,$$

where ϵ is some given positive number. If p is any positive integer the function

$$z^p (z + 1)^{-p} f(z)$$

will be bounded by ϵ at the points $z = \lambda_n$ for $n \geqq m$. If p is made sufficiently large this function will also be bounded by ϵ at the points $z = \lambda_n$ for $n < m$. Choosing p sufficiently large, the function

$$g(z) = \epsilon^{-1} z^p (z + 1)^{-p} f(z)$$

is bounded by 1 at the points $z = \lambda_n$ for all n in the range $n > \Gamma$. Since, clearly,

$$| g(z) | \leqq A \epsilon^{-1} e^{k|z|},$$

Corollary I states that $g(x)$ is bounded by $N(\Gamma, \gamma, k)$ for all large x. Then, expressing $f(z)$ is terms of $g(z)$,

$$f(z) = \epsilon (z + 1)^p z^{-p} g(z),$$

we see that for all sufficiently large x,

$$| f(x) | < 2\epsilon N(\Gamma, \gamma, k).$$

10. Proof of Theorem I. If $f(z)$ satisfies the conditions of Theorem I then $f(z)$ and $f(-z)$ each satisfy the conditions of Corollary I, from which we infer that $f(z)$ is bounded by some constant on the real axis. If we use the second part of Lemma I we see that

$$| f(z) | \leqq C e^{k|y|} \tag{32}$$

where C is some constant. It is then only necessary to show that the constant C can be chosen to depend exclusively on Γ, γ, and k.

From inequality (32) we see that if m is any positive integer the function

$$g_m(z) = f(z - m)$$

satisfies the inequality (11) with $A = C$. It is also bounded at a sequence of points $\{\mu_n\}$ which is obtained by translating the sequence $\{\lambda_n\}$ a distance m to the right. Corollary I therefore shows that

$$|g_m(x)| \leqq N(\Gamma, \gamma, k), \qquad x \geqq C^2/(\pi - k).$$

This in turn shows that $f(x) = g_m(x + m)$ is bounded by $N(\Gamma, \gamma, k)$ in the range $x \geqq -m + C^2/(\pi - k)$. Since m is an arbitrary positive integer it follows that $f(z)$ is bounded by $N(\Gamma, \gamma, k)$ on the entire real axis. If we again apply the second part of Lemma I to the function $f(z)$ we obtain the inequality (5).

11. In the second part of the present paper we shall be concerned with functions which are known to be bounded at all integer points save possibly those of a certain subsequence. Let $\{\mu_n\}$, $(n = 0, \pm 1, \cdots)$, be a sequence of positive and negative integers such that

(33) $$\mu_{n+1} - \mu_n > 0$$

and

(34) $$\mu_{n+1} - \mu_n \to \infty, \qquad n \to \pm \infty.$$

THEOREM III. *Let* $\{\mu_n\}$ *be a sequence of integers which satisfies* (33) *and* (34). *If* $f(z)$ *is an entire function which satisfies*

$$|f(z)| \leqq Ae^{k|z|}, \qquad k < \pi,$$

$$|f(m)| \leqq 1, \qquad m \notin \{\mu_n\},$$

then $f(z)$ *is uniformly bounded on the real axis, and*

(35) $$|f(z)| \leqq Ee^{k|y|}$$

If, in addition, $f(m) = o(1)$ *as* $m \to \pm \infty$, $m \notin \{\mu_n\}$, *then* $f(x) = o(1)$ *as* $x \to \pm \infty$.

Proof. The function $f(\rho z)$ where ρ is chosen in the range

$$1 < \rho < \pi/k$$

is bounded at the points

$$z_m = m/\rho, \qquad m \notin \{\mu_n\}.$$

Since ρ is chosen greater than 1 it is clear that there exists a subsequence $\{\lambda_n\}$ of the points z_m such that

$$|\lambda_n - n| \leqq \Gamma, \qquad (n = 0, \pm 1, \cdots),$$

where Γ is some constant. Also $|\lambda_n - \lambda_m| \geqq 1/\rho > 0$ when $n \neq m$. Since ρ is chosen less than π/k the function $f(\rho z)$ is of exponential type $k\rho < \pi$. Theorem I now shows that the function $f(\rho z)$, and hence $f(z)$, is bounded on the real axis. Inequality (35) then follows from Lemma I.

If $f(m) = o(1)$ as $m \to \pm\infty$, $m \notin \{\mu_n\}$, then by applying Corollary II to the functions $f(-\rho z)$ and $f(\rho z)$ we see that $f(x) = o(1)$ as $x \to \pm\infty$.

PART II.

1. A well known theorem of Szegö[6] states the following.

Let $f(z) = \sum_0^\infty a_n z^n$ *where the coefficients* a_n *take only a finite number of different values. If* $f(z)$ *is continuable beyond the unit circle then it is a rational function.*

We shall prove the following generalization of this theorem:

THEOREM I. *Let* $f(z) = \sum_0^\infty a_n z^n$ *where the coefficients* a_n *take only a finite number of different values. If* $f(z)$ *is bounded in a sector of the unit circle then it is a rational function.*

2. Suppose now that

$$f(z) = \sum_0^\infty a_n z^n$$

where the series converges in the unit circle $|z| < 1$ and the coefficients a_n are integers. It is known that if such a function is continuable beyond the unit circle then it is a rational function. This is a theorem which was conjectured by Pólya and was first proved by Carlson.[7] In view of the obvious similarity of these two theorems it is natural to ask if there is a corresponding generalization of the Pólya-Carlson theorem. That is, if $f(z)$ is bounded in a sector of the unit circle is it necessarily a rational function? The answer to this question is no, as is shown by the example

$$f(z) = \sum_1^\infty (1-z)^n z^{n!}.$$

This function is regular in the unit circle and its Taylor's expansion

---

[6] G. Szegö, "Tschebyscheffsche Polynome und nichfortsetzbare Potenzreihen," *Mathematische Annalen*, vol. 87 (1922), pp. 90-111.

[7] F. Carlson, "Über Potenzreihen mit ganzzahligen Koeffizienten, *Mathematische Zeitschrift*, vol. 9 (1921), pp. 1-13.

$$f(z) = \sum_0^\infty a_n z^{\lambda_n}$$

has integer coefficients. Moreover $f(z)$ is bounded in the sector of the unit circle $|\operatorname{amp} z| \leqq \alpha < \pi/3$. But $\lambda_n/n \to \infty$ so the Fabry gap theorem shows that it has the unit circle for a natural boundary and, in particular, is not a rational function.

3. Let $f(z)$ satisfy the conditions of Theorem I and suppose that $f(re^{i\theta})$ is bounded in the sector $\alpha < \theta < \beta$, $0 \leqq r < 1$. It is sufficient to consider the case in which $f(0) = 0$. Since $f(ze^{ip/q})$ will have only a finite number of different coefficients if p and q are integers, we also suppose that $\alpha < \pi < \beta$. We choose an $\epsilon > 0$ such that $\pi - \epsilon > \alpha$ and $\pi + \epsilon < \beta$, and then a $\delta > 0$ such that

$$\epsilon + \log(1 - \delta) > 0. \tag{1}$$

Let Γ be an open contour consisting of the circular arc $|\omega| = 1 - \delta$, $|\operatorname{amp}\omega| \leqq \pi - \epsilon$ and the two radial segments $1 - \delta \leqq |\omega| \leqq 1$, $\operatorname{amp}\omega = \pm(\pi - \epsilon)$. Then for any r in the range $1 - \delta < r < 1$,

$$a_n = \frac{1}{2\pi i}\int_{\Gamma_r} f(\omega)\omega^{-n-1}d\omega + \frac{1}{2\pi}\int_{\pi-\epsilon}^{\pi+\epsilon} f(re^{i\theta})r^{-n}e^{-in\theta}d\theta$$

where Γ_r denotes the portion of Γ that lies in the circle $|\omega| \leqq r$. This relation holds for negative n if we define a_n as 0 for $n = -1, -2, -3, \cdots$.

The last integral may be written in the form

$$\frac{1}{2\pi i}\int_{\pi-\epsilon}^{\pi+\epsilon} r^{-n}e^{-in\theta}\frac{dF(re^{i\theta})}{d\theta}d\theta = \frac{1}{2\pi i}\int_{\pi-\epsilon}^{\pi+\epsilon} r^{-n}e^{-in\theta}dF(re^{i\theta})$$

where

$$F(\omega) = \int_0^\omega f(t)t^{-1}dt.$$

Since $f(\omega)$ is bounded inside Γ it is clear that $F(\omega)$ satisfies a uniform Lipschitz condition inside Γ, $|F(\omega_1) - F(\omega_2)| \leqq M|\omega_1 - \omega_2|$ when ω_1 and ω_2 lie inside Γ. Thus, as $r \to 1$, $F(re^{i\theta})$ tends uniformly to a limit $F(e^{i\theta})$ in the range $\pi - \epsilon \leqq \theta \leqq \pi + \epsilon$. Letting r appproach 1 we obtain

$$a_n = \frac{1}{2\pi i}\int_\Gamma f(\omega)\omega^{-n-1}d\omega + \frac{1}{2\pi i}\int_{\pi-\epsilon}^{\pi+\epsilon} e^{-in\theta}dF(e^{i\theta}). \tag{2}$$

Since $F(e^{i\theta})$ satisfies a uniform Lipschitz condition the last integral is $o(1)$ as $n \to \pm\infty$.

Let

$$\phi(z) = \frac{1}{2\pi i}\int_\Gamma f(\omega)\omega^{-1}e^{-z\log\omega}d\omega \tag{3}$$

where $\log \omega$ has its principal value. For integer values of z we infer from (2) that

$$\phi(n) = a_n + o(1), \qquad n \to \pm \infty. \tag{4}$$

Furthermore, $f(\omega)$ is bounded on Γ, and, from (1),

$$|\log \omega| \leqq \pi - \epsilon - \log(1 - \delta) < \pi, \qquad \omega \in \Gamma;$$

hence we infer that $\phi(z)$ is an entire function of exponential type less than π,

$$|\phi(z)| \leqq A_1 e^{k|z|}, \qquad k < \pi.$$

Now (4) implies that $\phi(z)$ is bounded at the integer points, so from Theorem I of Part I we obtain the inequality

$$|\phi(z)| \leqq A_2 e^{k|y|}. \tag{5}$$

This inequality shows that if $\{\mu_n\}$ is any sequence of real numbers the functions $\phi(z - \mu_n)$ will be uniformly bounded in every finite region, and so, by Ascoli's convergence theorem concerning equicontinuous functions, there will be a subsequence which converges to an entire function which satisfies (5). The family of functions $\phi(z - \mu_n)$, μ_n real, is a *normal family.*

4. Let $b_1, b_2, \cdots, b_m$ be the different values that a_n can have,

$$|b_i - b_j| \geqq b > 0, \qquad i \neq j.$$

If r is a given positive integer there will be positive integers $p = p(r)$ and $q = q(r)$, $p > q \geqq 0$, such that

$$a_{p+\nu} = a_{q+\nu}, \qquad 0 \leqq \nu \leqq r.$$

Then either

$$a_{p+\nu} = a_{q+\nu}, \qquad \nu \geqq 0 \tag{i}$$

or

(ii) there is an integer $R > r$ such that

$$\begin{aligned} a_{p+\nu} &= a_{q+\nu}, \qquad 0 \leqq \nu \leqq R - 1 \\ a_{p+R} &\neq a_{q+R}, \end{aligned} \tag{6}$$

Suppose that there are arbitrarily large integers r for which case (ii) is true. It is to be shown that this leads to a contradiction. Let

$$g_r(z) = \phi(z + p + R) - \phi(z + q + R)$$

where p, q, and R are functions of r which satisfy (6) and $p > q \geqq 0$. Then $g_r(z)$ is an entire function and, in virtue of (5),

$$|g_r(z)| \leqq 2A_2 e^{k|y|}.$$

As r becomes large R does also, so relation (4) shows that

$$(7) \qquad g_r(m) = a_{m+p+R} - a_{m+q+R} + o(1), \qquad r \to \infty,$$

for each fixed integer m. But as r becomes large some subsequence of the functions $g_r(z)$ tends to a limit $g(z)$ which is an entire function and satisfies

$$\text{(a)} \qquad |g(z)| \leqq 2A_2 e^{k|y|}, \qquad k < \pi,$$

$$\text{(b)} \qquad |g(0)| \geqq b > 0,$$

$$\text{(c)} \qquad g(m) = 0, \qquad (m = -1, -2, \cdots).$$

Properties (b) and (c) are obtained from (6) and (7). But a theorem of Carlson states that a function which satisfies (a) and (c) must vanish identically. This contradicts (b).

It has thus been shown that there is an integer for which case (i) is true, and this clearly implies that $f(z)$ is a rational function of the form

$$f(z) = P(z)/(1 - z^{p-q})$$

where $P(z)$ is a polynomial.

5. The methods of the last paragraphs can be used to obtain a theorem concerning power series in which all the coefficients are bounded save possibly those of a Faber sequence. The following theorem is allied to a result of Paley and Wiener,[8] although it does not contain their theorem, nor is it contained in their theorem (for a more general result cf. **6**).

THEOREM II. *Let* $f(z) = \sum_{0}^{\infty} a_n z^n$ *be bounded in a sector of the unit circle. Let* $\{n_\kappa\}$ *be a sequence of integers such that*

$$n_{\kappa+1} - n_\kappa \to \infty, \qquad \kappa \to \infty.$$

If the coefficients a_n *are bounded, save possibly when* $n \in \{n_\kappa\}$, *then in fact all coefficients are bounded. If* $a_n = o(1)$ *as* $n \to \infty$, $n \notin \{n_\kappa\}$, *then all coefficients are* $o(1)$ *as* $n \to \infty$.

If $f(z)$ satisfies the conditions of Theorem II it can be written in the form

$$f(z) = f_1(z) + f_2(z)$$

where $f_1(z)$ has bounded coefficients, and

---

[8] Paley and Wiener, *Fourier Transforms in the Complex Domain*, p. 124.

$$f_2(z) = \sum_{n \epsilon \{n_\kappa\}} a_n z^n.$$

Now $f_1(z)$ is regular in $|z| < 1$, and $f_2(z)$ is not continuable beyond its circle of convergence according to well known gap theorems. Since $f(z)$ and $f_1(z)$, and therefore $f_2(z)$, are regular in a sector of the unit circle, $f_2(z)$ must be regular in the unit circle; hence $f(z)$ is regular in $|z| < 1$.

By considering the function $f(ze^{i\theta}) - f(0)$ we suppose, without loss of generality, that $f(0) = 0$ and that $f(z)$ is bounded in a sector which includes the point $z = -1$ and is symmetric about the real axis. Let the contour Γ be the same as in **3** and let $\phi(z)$ be defined as in (3). Then $\phi(z)$ satisfies the relation (4) and so it is bounded at all integer points except possibly those of the sequence $\{n_\kappa\}$. Also

$$\phi(z) \leqq A_3 e^{k|z|}, \qquad k < \pi.$$

From Theorem III of Part I we immediately infer that $\phi(z)$ is bounded on the entire real axis, and then, as relation (4) shows, $a_n = O(1)$ as $n \to \infty$. The last statement of the theorem is proved in the same way, using Theorem III of Part I. We remark that the condition $n_{\kappa+1} - n_\kappa \to \infty$ as $n \to \infty$ in Theorem II cannot be replaced by the more general condition $n_\kappa/\kappa \to \infty$ as $\kappa \to \infty$. This is shown by the example $f(z) = \sum_1^\infty (1 - z)^n z^{n!}$, which was mentioned in connection with the Pólya-Carlson theorem.

6. Finally, we remark that the condition in Theorems I and II that $f(z)$ is bounded in a sector of the unit circle is unnecessarily restrictive. It would be sufficient, for example, to suppose that $|f(re^{i\theta})| \leqq q(\theta)$ in a sector $\alpha < \theta < \beta$, $0 \leqq r < 1$, where $q(\theta)$ is some function which is Lebesgue integrable over (α, β). With this hypothesis we would again need only to consider the case in which $\alpha < \pi < \beta$. We would then choose an $\epsilon > 0$ such that $\pi - \epsilon > \alpha$, $\pi + \epsilon < \beta$, and also such that $q(\pi - \epsilon) + q(\pi + \epsilon) < \infty$. In this case the function $F(re^{i\theta})$, $F(\omega) = \int_0^\omega f(t) t^{-1} dt$, would converge almost everywhere in $(\pi - \epsilon, \pi + \epsilon)$ to an absolutely continuous function as $r \to 1$. The subsequent proof would be essentially unchanged.

We also note that it is possible to prove Theorem I in another way; namely, by combining the method of Szegö with a method of Dienes.[9]

UNIVERSITY OF ILLINOIS,
STANFORD UNIVERSITY.

---

[9] P. Dienes, *The Taylor Series*, Oxford (1931), p. 504.

NOTE CONCERNING THE CONFORMAL AND EQUILONG GEOMETRIES OF FOURTH AND FIFTH ORDER HORN SETS.*

By Mary Elizabeth Ladue.

The existence of a unique nontrivial equilong invariant and of a unique nontrivial conformal invariant has been proved for each horn angle of a given order of contact and the actual formulae for several of these invariants have been found.[1] The theorem stated below gives, in terms of the 1943 paper, two previously unpublished formulae for the equilong measures of fourth and fifth order horn angles.

Theorem I. *A horn angle of fourth order has the expression*

$$(1)\qquad \frac{(\Delta a_5)^5}{(\Delta a_5)^4 - 3\Delta a_5(\Delta a_6)^2\Delta a_7 + (\Delta a_5)^2[2\Delta a_6\Delta a_8 + (\Delta a_7)^2] - (\Delta a_5)^3\Delta a_9}$$

as its unique equilong invariant and the fifth order horn angle has

$$(2)\qquad \frac{(\Delta a_6)^6}{\begin{aligned}(\Delta a_7)^5 - 4\Delta a_6(\Delta a_7)^3\Delta a_8 + 3(\Delta a_6)^2[(\Delta a_7)^2\Delta a_9 + \Delta a_7(\Delta a_8)^2]\\ - 2(\Delta a_6)^3(\Delta a_7\Delta a_{10} + \Delta a_8\Delta a_9) + (\Delta a_6)^4\Delta a_{11}\end{aligned}}$$

as its unique equilong invariant.

Associated with each n-th order horn set is a space of $n+1$ dimensions, called the conformal K_{n+1}-space, in which the distance between two points is given by the conformal measure of the horn angle associated with that pair of points and the fundamental group is the set of transformations, induced by conformal transformations of curves of the horn set, which leave invariant this distance. The equations of these induced transformations for the conformal K_5-space and the conformal K_6-space are given by equations (3) and (4);

---

* Received April 24, 1944; Revised September 9, 1944.

[1] See the author's previous paper, "Conformal geometry of Horn angles of higher order," *American Journal of Mathematics*, vol. 65 (1943), pp. 455-476, for references, definitions of terms and description of methods used in this note.

$$
\begin{aligned}
&V = m_1{}^4 v + h_1,\\
&W = m_1{}^5 w + 3m_1{}^3 m_2 v + h_2,\\
&X = m_1{}^6 x + 4m_1{}^4 m_2 w + 2m_1{}^2 m_3 v + h_3,\\
(3)\qquad &Y = m_1{}^7 y + 5m_1{}^5 m_2 x + m_1{}^3(3m_3 + m_2{}^2) w + m_1 m_4 v + h_4,\\
&Z = m_1{}^8 z + 6m_1{}^6 m_2 y + m_1{}^4(4m_3 + m_2{}^2) x\\
&\qquad + 2m_1{}^2(m_4 - m_2 m_3 + 3m_2{}^3) w\\
&\qquad + m_1{}^4(6m_2 m_4 + 4m_3{}^2 - 54m_2{}^2 m_3 + 81m_2{}^4) v + h_5;
\end{aligned}
$$

$$
\begin{aligned}
&U = m_1{}^5 u + h_1,\\
&V = m_1{}^6 v + 4m_1{}^4 m_2 u + h^2,\\
&W = m_1{}^7 w + 5m_1{}^5 m_2 v + 3m_1{}^3 m_3 u + h_3,\\
(4)\qquad &X = m_1{}^8 x + 6m_1{}^6 m_2 w + m_1{}^4(4m_3 + \tfrac{5}{3} m_2{}^2) v + 2m_1{}^2 m_4 u + h_4,\\
&Y = m_1{}^9 y + 7m_1{}^7 m_2 x + m_1{}^5(5m_3 + \tfrac{13}{3} m_2{}^2) w\\
&\qquad + m_1{}^3(3m_4 + m_2 m_3 + \tfrac{5}{3} m_2{}^3) v + m_1 m_5 u + h_5,\\
&Z = m_1{}^{10} z + 8m_1{}^8 m_2 y + m_1{}^6(6m_3 + 8m_2{}^2) x\\
&\qquad + m_1{}^4(4m_4 + 4m_2 m_3 + \tfrac{8}{3} m_2{}^3) w\\
&\qquad + m_1{}^2(2m_5 - 4m_2 m_4 - 3m_3{}^2 + 42m_2{}^2 m_3 - \tfrac{200}{3} m_2{}^4) v\\
&\qquad + (8m_2 m_5 + 12m_3 m_4 - 96m_2{}^2 m_4\\
&\qquad\qquad - 108m_2 m_3{}^2 + 768m_2{}^2 m_3 - 1024m_2{}^5) u + h_6.
\end{aligned}
$$

THEOREM II. *The fundamental transformation and invariant metric of the geometry of the conformal space $K_5(\gamma, \gamma', \gamma'')$ are given by equations* (1) *and* (3) *respectively; those of the geometry of the conformal space $K_6(\gamma, \gamma', \gamma'', \gamma''')$ by* (2) *and* (4).

Also associated with each n-th order horn set is another $n + 1$ dimensional space, called the equilong K'_{n+1}-space, in which the distance is given by the equilong measure and the fundamental transformation is induced by the equilong transformation of the horn set. It has been shown by Kasner and the author that the equilong K'-spaces associated with horn sets of the first, second, and third orders are identical with the corresponding conformal K-spaces of those horn sets. Computation of the equilong K'_5-space and the equilong K'_6-space shows, upon comparison with the results stated in Theorem II, that this analogy holds also for the conformal and equilong K-spaces associated with fourth and fifth order horn sets.

THEOREM III. *The geometries of the equilong and conformal K-spaces are identical for horn sets of orders one to five.*

BARNARD COLLEGE,
COLUMBIA UNIVERSITY.

GEOMETRY OF SCALE CURVES IN CONFORMAL MAPS.*

By Edward Kasner and John De Cicco.

1. General summary. Let a surface Σ be mapped in any arbitrary point-to-point fashion (continuous and differentiable) upon a plane Π with cartesian coordinates (x, y). The scale function $\sigma = ds/dS$ which is the ratio of the corresponding elements of arc length in the plane Π and in the surface Σ depends, in general, not only upon the point but also upon the direction. It is independent of the direction if, and only if, the mapping of Σ upon Π is conformal.

We wish to present the fundamental theorems concerning the scale function σ in a general conformal mapping of a surface Σ upon a plane Π. A scale curve is the locus of a point along which the scale function σ does not vary. Therefore, in the conformal case, there are ∞^1 scale curves, defined by the finite equation $\sigma(x, y) = \text{const.}$ We shall be chiefly concerned with these scale curves. In non-conformal maps, which we study in other papers, σ is a function of x, y, y' and so we have ∞^2 scale curves. The application to the cartography of the sphere is given in the final section.

In our work, we shall obtain new characterizations of surfaces applicable upon surfaces of revolution. In his study of the geodesics of a surface, Kasner has obtained various properties of these surfaces applicable upon surfaces of revolution.[1] If a surface possesses an isothermal system of geodesics, then it is applicable upon a surface of revolution, the geodesics corresponding to meridians. The only surfaces upon which more than one isothermal system of geodesics can exist are the surfaces of constant gaussian curvature.[2] Kasner has also proved that the only surfaces which can be represented point by point upon a plane so that the geodesics are represented by parabolas are those of constant curvature and also special types of geodesically equivalent surfaces applicable upon certain surfaces of revolution.[3]

---

* Received July 1, 1944. Presented to the American Mathematical Society, February, 1944. See two related papers in *Proceedings of the National Academy of Sciences*, 1944.

[1] Kasner, "Isothermal systems of geodesics," *Transactions of the American Mathematical Society*, vol. 5 (1904), pp. 55-60.

[2] For a recent discussion of these theorems, see the paper by De Cicco, "New proofs of the theorems of Beltrami and Kasner on linear families," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 407-412.

[3] Kasner, "Surfaces whose geodesics may be represented in the plane by parabolas," *Transactions of the American Mathematical Society*, vol. 6 (1905), pp. 141-158.

Any family of ∞^1 curves can represent the scale curves of a conformal map of any one of a certain class of surfaces Σ upon a plane Π. If it be demanded that the gaussian curvature G be constant along the scale curves, it is found that the set of scale curves must belong to a certain special class of families of curves. That is, if $f(x, y) = \text{const.}$, represents the scale curves of a conformal map of a surface Σ upon a plane Π such that the gaussian curvature G is constant along each of the scale curves, then the function f must satisfy a certain partial differential equation of the fourth order. If our family of scale curves is defined as the integral curves of a differential equation of the first order $y' = \tan\theta(x, y)$, where θ is the inclination of the direction through the point (x, y), then the function θ must satisfy a certain partial differential equation of the third order. We shall term such a family of ∞^1 curves a family of the type λ or a *quasi-isothermal family.*

Any isothermal family of curves is of the type λ; but not every such family of curves is isothermal. Thus our new quasi-isothermal families of curves may be considered to be wide generalizations of isothermal families. For example, the family of similar ellipses $x^2 + 2y^2 = \text{const.}$, is quasi-isothermal but not isothermal.

We prove that, if an isothermal family represents the scale curves of a conformal map of a surface Σ upon a plane Π such that the gaussian curvature is constant along each of these scales, then either Σ is developable or it is a surface applicable upon a general surface of revolution. In the case of a developable surface, the scale curves can be any isothermal family; but in the case of a surface applicable upon a surface of revolution, the scale curves are either ∞^1 parallel straight lines or ∞^1 concentric circles. In the latter case, the parallels of Σ correspond to the parallel straight scales or to the concentric circular scales.

Again suppose that the quasi-isothermal family of curves is a parallel family (the orthogonal trajectories of any system of ∞^1 straight lines), then they must be parallel straight lines or concentric circles. The surface Σ is either developable or applicable upon a surface of revolution.

The only simply-infinite families of straight lines or circles that are of the type λ are the pencils of straight lines or circles. The surface Σ must again be either developable or applicable upon a surface of revolution. In the latter case, the scale curves must be either parallel straight lines or concentric circles. This is a wide generalization of theorems of Lagrange.

The application of the preceding results to any conformal map of a sphere upon a plane yields *new characterizations of the Mercator, Ptolemy (stereographic) and Lambert conical projections. These are the only conformal maps* *of a sphere upon a plane such that the scale curves form an isothermal family or a parallel family.* The Mercator projection is the only conformal map of a sphere upon a plane such that the scale curves are straight lines; whereas the stereographic and Lambert conical projections are the only conformal maps of a sphere upon a plane such that the scale curves are circles. These theorems are simple but the proofs are complicated.

2. Conformal maps of a surface Σ upon a plane Π. If the mapping of Σ upon Π is conformal, the square of the linear-element dS of Σ is of the form

$$dS^2 = e^{2\lambda(x,y)}(dx^2 + dy^2). \tag{1}$$

The scale function $\sigma = ds/dS$ is then given by $\sigma = e^{-\lambda(x,y)}$. The gaussian curvature G of the surface Σ is

$$G = - e^{-2\lambda(x,y)}(\lambda_{xx} + \lambda_{yy}). \tag{2}$$

The geodesic curvature K of a curve upon the surface Σ is then

$$K = \frac{y'' - (1 + y'^2)(\lambda_y - y'\lambda_x)}{e^{\lambda}(1 + y'^2)^{3/2}}. \tag{3}$$

Thus if K and k are the corresponding curvatures in the surface Σ and in the plane Π, then

$$K = e^{-\lambda}\left[k - \frac{\lambda_y - y'\lambda_x}{(1 + y'^2)^{3/2}}\right]. \tag{4}$$

3. Scale curves in conformal maps. The scale curves of the conformal map (1) are given by $\lambda(x, y) = \text{const.}$ By (3) and (4), the curvatures of the scale curves on the surface and plane are related by

$$K = e^{-\lambda}[k - (\lambda_x^2 + \lambda_y^2)^{1/2}], \tag{5}$$

where

$$k = - \frac{\lambda_y^2\lambda_{xx} - 2\lambda_x\lambda_y\lambda_{xy} + \lambda_x^2\lambda_{yy}}{(\lambda_x^2 + \lambda_y^2)^{3/2}}. \tag{6}$$

By (5) and (6), it is found that if the scale curves are straight lines in Π and geodesics in Σ, then λ is constant. This means that Σ is developable and the conformal map (1) is an unrolling of the surface Σ upon a plane which in turn is mapped by a similitude upon the plane Π. Henceforth we shall exclude any such map from consideration, so that λ is any non-constant function of (x, y).

Suppose now that $f(x,y) = \text{const.}$, is any simple family of curves. These can represent the scale curves of that class of surfaces for which $\lambda = \lambda(f)$. The gaussian curvature G of this class of surfaces Σ is

$$G = -e^{-2\lambda(f)}[\lambda_{ff}(f_x^2 + f_y^2) + \lambda_f(f_{xx} + f_{yy})]. \tag{7}$$

4. Families of curves of the type λ (quasi-isothermal). We shall say that any simple family of curves $f(x,y) = \text{const.}$, is of the type λ or is quasi-isothermal if it represents the scale curves of a conformal map of a surface Σ upon a plane Π such that the gaussian curvature G is constant along each of the scale curves. Not any simple family of curves can satisfy these requirements. In this section, we shall derive the condition that a simple family of curves shall be of the type λ.

For this purpose, it is found convenient to introduce the linear operator

$$\mathfrak{D} = f_y\partial/\partial x - f_x\partial/\partial y. \tag{8}$$

It is seen that $\mathfrak{D}$ will annihilate only those functions which depend upon f alone.

Now if the gaussian curvature G is constant along each of the scale curves $f(x,y) = \text{const.}$, it follows that G must be a function of f only. Therefore using the linear operator $\mathfrak{D}$ upon the equation (7), we discover that

$$\frac{\mathfrak{D}(f_{xx} + f_{yy})}{\mathfrak{D}(f_x^2 + f_y^2)} = -\frac{\lambda_{ff}}{\lambda_f}. \tag{9}$$

Using the linear operator $\mathfrak{D}$ upon this equation, we obtain the following result.

THEOREM 1. *The simple family of curves $f(x,y) = \text{const.}$ is of the type λ if and only if the function f satisfies the partial differential equation of the fourth order*

$$\mathfrak{D}\,\frac{\mathfrak{D}(f_{xx} + f_{yy})}{\mathfrak{D}(f_x^2 + f_y^2)} = 0. \tag{10}$$

If $\lambda(x,y)$ obeys an equation of the form $\lambda_{xx} + \lambda_{yy} = \phi(\lambda)$, then $\lambda(x,y) = c$ is a quasi-isothermal family, and c is then called the *quasi-isothermal parameter.*

It is found that if $y' = \tan\theta(x,y)$ is the differential equation of the first order defining any family of curves of the type λ, then the inclination θ must satisfy the partial differential equation of the third order

$$\begin{aligned}(\theta_x\tan\theta - \theta_y)(\partial/\partial x + \tan\theta\,\partial/\partial y)\log(\theta_{xx} + \theta_{yy})\\ + \tan\theta(\theta_{yy} - \theta_{xx} - 4\theta_x\theta_y) + (1 - \tan^2\theta)(\theta_{xy} - \theta_x^2 + \theta_y^2) = 0.\end{aligned} \tag{11}$$

5. Isothermal and parallel families of the type λ. The class of quasi-isothermal families contains the isothermal families as a special subclass. According to a theorem of Lie, the simple family of curves $f(x,y) = \text{const.}$ is isothermal if and only if $\mathfrak{D}[(f_{xx}+f_{yy})/(f_x^2+f_y^2)] = 0$, or the simple family of curves defined by the differential equation of the first order $y' = \tan\theta(x,y)$ is isothermal if and only if $\theta_{xx}+\theta_{yy} = 0$. Placing these conditions into (10) or (11), we find that any isothermal family is of the type λ.

THEOREM 2. *If the ∞^1 scale curves of a conformal map of a surface Σ upon a plane Π form an isothermal system such that the gaussian curvature G of Σ is constant along each of the scale curves, then Σ is either developable or applicable upon a surface of revolution. In the former case, the scale curves may be any isothermal family, but in the latter case the scale curves must be either parallel straight lines or concentric circles.*

If $f(x,y) = \text{const.}$, is an isothermal family, we may so choose the parameter that f satisfies the Laplace equation $f_{xx}+f_{yy} = 0$. By (9), it then follows that

$$\lambda_{ff}\mathfrak{D}(f_x^2+f_y^2) = 0. \tag{12}$$

If $\lambda_{ff} = 0$, then λ is linear in the harmonic function f only. By (1) and (2), it follows that the surface Σ is developable and $f(x,y) = \text{const.}$, can represent any isothermal family.

Consider next the condition $\lambda_{ff} \neq 0$. By the preceding equation and the condition that f is harmonic, we must have

$$\mathfrak{D}(f_x^2+f_y^2) = 0, \qquad f_{xx}+f_{yy} = 0. \tag{13}$$

We have to solve these equations simultaneously.

Introduce the linear operators

$$\begin{aligned} \partial/\partial u &= \tfrac{1}{2}(\partial/\partial x - i\,\partial/\partial y), & \partial/\partial v &= \tfrac{1}{2}(\partial/\partial x + i\,\partial/\partial y), \\ \partial/\partial x &= (\partial/\partial u + \partial/\partial v), & \partial/\partial y &= i(\partial/\partial u - \partial/\partial v). \end{aligned} \tag{14}$$

Kasner has termed the operator $\partial/\partial u$ the mean derivative and the operator $\partial/\partial v$ the phase derivative. (These operators are important in the development of the geometry of polygenic functions.[4])

---

[4] Kasner, "The second derivative of a polygenic function," *Transactions of the American Mathematical Society*, vol. 30 (1928), pp. 803-818. Also Kasner and De Cicco. "The derivative circular congruence-representation of a polygenic function," *American Journal of Mathematics*, vol. 61 (1939), pp. 995-1003. See a paper appearing in *Scripta Mathematica*, 1945.

The equations (13) then reduce to

$$f_v^2 f_{uu} = f_u^2 f_{vv}, \qquad f_{uv} = 0, \tag{15}$$

where subscripts denote the operators (14), which behave formally like partial derivatives. These equations show that all the real solutions are equivalent by a similitude to

$$f = n \log(x^2 + y^2), \quad \text{or} \quad f = x, \tag{16}$$

where $n \neq 0$.

Therefore the surface must be applicable upon a surface of revolution. In the first case the scale curves are concentric circles and in the second case the scale curves are parallel straight lines. The corresponding linear-elements of these two types are

$$dS^2 = e^{2\lambda(x^2+y^2)}(dx^2 + dy^2), \quad \text{or} \quad dS^2 = e^{2\lambda(x)}(dx^2 + dy^2). \tag{17}$$

These are conformally equivalent by the transformation $U = \log u^n$.

(We note that in the imaginary domain, the complete solutions of (15) will also include the cases where f is monogenic in $u = x + iy$, or monogenic in $v = x - iy$. In either of these two cases, the corresponding surfaces are imaginary and are applicable to imaginary surfaces of revolution. The scale curves then are minimal lines of the same kind).

THEOREM 3. *If the ∞^1 scale curves of a conformal map of a surface Σ upon a plane Π form a parallel family such that the gaussian curvature G is constant along each of the scale curves, then the scale curves are either parallel straight lines or concentric circles, and Σ is developable or applicable upon a surface of revolution.*

The condition for a parallel family of curves is $\mathfrak{D}(f_x^2 + f_y^2) = 0$. By equation (9), it follows that $\mathfrak{D}(f_{xx} + f_{yy}) = 0$. Hence from the proof of Theorem 2, it follows that the scale curves must be parallel straight lines or concentric circles. The surface Σ must then be either developable or applicable upon a surface of revolution.

6. Simply-infinite families of straight lines which are of the type λ (quasi-isothermal). We shall discuss the following result:

THEOREM 4. *The only families of ∞ straight lines which are of the type λ are the pencils of straight lines.*

Consider the ∞^1 straight lines

$$y = x\alpha(f) + \beta(f), \tag{18}$$

where α and β are certain functions of f alone. This defines f as a certain function of (x, y). Differentiating the above partially with respect to x and y, we obtain

$$\begin{aligned} f_x^2 + f_y^2 &= \frac{1+\alpha^2}{(x\alpha_f + \beta_f)^2}, \\ f_{xx} + f_{yy} &= \frac{2\alpha\alpha_f}{(x\alpha_f + \beta_f)^2} - \frac{(1+\alpha^2)(x\alpha_{ff} + \beta_{ff})}{(x\alpha_f + \beta_f)^3}. \end{aligned} \tag{19}$$

Using the operator $\mathfrak{D}$ upon the preceding two equations, we find

$$\begin{aligned} \mathfrak{D}(f_x^2 + f_y^2) &= -\frac{2(1+\alpha^2)\alpha_f}{(x\alpha_f + \beta_f)^4}, \\ \mathfrak{D}(f_{xx} + f_{yy}) &= -\frac{4\alpha_f^2 + (1+\alpha^2)\alpha_{ff}}{(x\alpha_f + \beta_f)^4} + \frac{3(1+\alpha^2)\alpha_f(x\alpha_{ff} + \beta_{ff})}{(x\alpha_f + \beta_f)^5}. \end{aligned} \tag{20}$$

Now if $\alpha_f = 0$, our straight lines are parallel. Assuming that $\alpha_f \neq 0$, it follows from (20) that

$$\frac{\mathfrak{D}(f_{xx} + f_{yy})}{\mathfrak{D}(f_x^2 + f_y^2)} = \frac{2\alpha\alpha_f}{1+\alpha^2} + \frac{\alpha_{ff}}{2\alpha_f} - \frac{3(x\alpha_{ff} + \beta_{ff})}{2(x\alpha_f + \beta_f)}. \tag{21}$$

Applying the linear operator $\mathfrak{D}$ upon this, we find

$$\mathfrak{D}\,\frac{\mathfrak{D}(f_{xx} + f_{yy})}{\mathfrak{D}(f_x^2 + f_y^2)} = \frac{-3(\alpha_{ff}\beta_f - \alpha_f\beta_{ff})}{2(x\alpha_f + \beta_f)^3}. \tag{22}$$

By the condition (10) for quasi-isothermal systems of curves and the preceding equation, it follows that the ∞^1 straight lines must form a pencil. This completes the proof of Theorem 4.

7. Simply-infinite families of circles which are of the type λ (quasi-isothermal). Next we shall prove the following extension of the preceding result.

THEOREM 5. *The only families of ∞^1 circles which are of the type λ are the pencils of circles.*

Consider the family of ∞^1 circles

$$x^2 + y^2 = 2[xa(f) + yb(f) + c(f)], \tag{23}$$

where (a, b, c) are functions of f. This equation defines f as a function of

(x, y). We shall have occasion to use the radius r which is given by the formula

$$r^2 = a^2 + b^2 + 2c. \tag{24}$$

Differentiating (23) partially with respect to x and y, we find

$$\begin{aligned} f_x^2 + f_y^2 &= \frac{r^2}{(xa_f + yb_f + c_f)^2}, \\ f_{xx} + f_{yy} &= \frac{2rr_f}{(xa_f + yb_f + c_f)^2} - \frac{r^2(xa_{ff} + yb_{ff} + c_{ff})}{(xa_f + yb_f + c_f)^3}. \end{aligned} \tag{25}$$

Using the linear operator $\mathcal{D}$ upon these equations, we find

$$\begin{aligned} \mathcal{D}(f_x^2 + f_v^2) &= \frac{-2r^2[(y-b)a_f - (x-a)b_f]}{(xa_f + yb_f + c_f)^4}, \\ \mathcal{D}(f_{xx} + f_{yy}) &= -\frac{4rr_f[(y-b)a_f - (x-a)b_f] + r^2[(y-b)a_{ff} - (x-a)b_{ff}]}{(xa_f + yb_f + c_f)^4} \\ &\quad + \frac{3r^2(xa_{ff} + yb_{ff} + c_{ff})[(y-b)a_f - (x-a)b_f]}{(xa_f + yb_f + c_f)^5}. \end{aligned} \tag*{26)}$$

Now if $\mathcal{D}(f_x^2 + f_y^2) = 0$, it can be shown that the circles are concentric. Assuming that $\mathcal{D}(f_x^2 + f_y^2) \neq 0$, it follows that

$$\begin{aligned} \frac{\mathcal{D}(f_{xx} + f_{yy})}{\mathcal{D}(f_x^2 + f_y^2)} &= \frac{2r_f}{r} + \frac{[(y-b)a_{ff} - (x-a)b_{ff}]}{2[(y-b)a_f - (x-a)b_f]} \\ &\quad - \frac{3(xa_{ff} + yb_{ff} + c_{ff})}{2(xa_f + yb_f + c_f)}. \end{aligned} \tag{27}$$

From this we find

$$\begin{aligned} \mathcal{D}\frac{\mathcal{D}(f_{xx} + f_{yy})}{\mathcal{D}(f_x^2 + f_y^2)} &= \frac{r^2(a_{ff}b_f - a_fb_{ff})}{2[(y-b)a_f - (x-a)b_f]^2(xa_f + yb_f + c_f)} \\ &\quad - \frac{\begin{array}{l} 3[(a_{ff}b_f - a_fb_{ff})(ax + by + 2c) \\ \qquad + (b_fc_{ff} - b_{ff}c_f)(x-a) + (a_{ff}c_f - a_fc_{ff})(y-b)] \end{array}}{2(xa_f + yb_f + c_f)^3}. \end{aligned} \tag{28}$$

Upon setting the right hand side of this equation equal to zero, and using the parametric form of the ∞^1 circles: $x = a + r(1-t^2)/(1+t^2)$, $y = b + 2rt/(1+t^2)$, we find

$$\begin{aligned} &r(1+t^2)(a_{ff}b_f - a_fb_{ff})[(r_f - a_f)t^2 + 2b_ft + (r_f + a_f)]^2 \\ &- 3(b_ft^2 + 2a_ft - b_f)^2 \begin{bmatrix} (a_{ff}b_f - a_fb_{ff})\{(r-a)t^2 + 2bt + (r+a)\} \\ + (b_fc_{ff} - b_{ff}c_f)(1-t^2) + 2t(a_{ff}c_f - a_fc_{ff}) \end{bmatrix} = 0. \end{aligned} \tag*{29)}$$

It is obvious that if $a_{ff}b_f - a_f b_{ff} = 0$, then this identity in t yields the pencils of circles. We shall prove now that $a_{ff}b_f - a_f b_{ff}$ must be zero. For otherwise if $a_{ff}b_f - a_f b_{ff} \neq 0$, then $a_f \neq 0$. Thus we may take $f = a$ and $b_{aa} \neq 0$. The preceding equation reduces to

$$(30)\quad rb_{aa}(1+t^2)[(r_a-1)t^2+2b_a t+(r_a+1)]^2$$
$$+3(b_a t^2+2t-b_a)^2\begin{bmatrix}-b_{aa}\{(r-a)t^2+2bt+(r+a)\}\\+(b_a c_{aa}-b_{aa}c_a)(1-t^2)-2tc_{aa}\end{bmatrix}=0.$$

Setting the coefficient of t^6 equal to zero, we find

$$(31)\qquad rb_{aa}(r_a-1)^2 = 3b_a^2[b_{aa}(r-a)+(b_a c_{aa}-b_{aa}c_a)].$$

By means of this equation, the equation (30) may be written as

$$(32)\quad rb_{aa}(1+t^2)[(r_a-1)t^2+2b_a t+(r_a+1)]^2$$
$$=(b_a t^2+2t-b_a)[(rb_{aa}/b_a^2)(r_a-1)^2(t^2-1)$$
$$+6t(bb_{aa}+c_{aa})+6rb_{aa}].$$

Upon setting the coefficient of t^5 equal to zero, we find

$$(33)\quad 6(bb_{aa}+c_{aa}) = \frac{4rb_{aa}}{b_a}(r_a-1) - \frac{4rb_{aa}}{b_a^3}(r_a-1)^2.$$

Substituting this into (32) and simplifying, we obtain

$$(34)\quad (1+t^2)[(r_a-1)t^2+2b_a t+(r_a+1)]^2$$
$$=(b_a t^2+2t-b_a)^2[(1/b_a^2)(r_a-1)^2(t^2-1)$$
$$+(4t/b_a^3)(r_a-1)\{b_a^2-(r_a-1)\}+6].$$

Upon setting the coefficient of t equal to zero, we discover that $b_a = 0$. This is impossible. Hence the only families of ∞^1 circles which are of the type λ are the pencils of circles. This completes the proof of Theorem 5.

By Theorems 4 and 5, it follows that if the ∞^1 scale curves of a conformal map of a surface Σ upon a plane Π are straight lines or circles such that the gaussian curvature G is constant along each of the scale curves, then Σ is developable or applicable upon a surface of revolution. In the latter case, the straight scales must be parallel, or the circular scales must be concentric.

8. Application of the preceding results to the sphere. The two important types of conformal maps (17) upon a plane Π of a surface Σ applicable upon a surface of revolution are reducible under the conformal transformation $U = \log u^n$ to the single case

$$dS^2 = e^{2\lambda(x)}(dx^2 + dy^2). \tag{35}$$

The gaussian curvature G is then

$$G = - e^{-2\lambda(x)}\lambda_{xx}. \tag{36}$$

Let us now suppose that the surface Σ is a sphere so that the gaussian curvature G is a positive constant. By the application of an appropriate magnification, G may be reduced to unity, that is, $G = 1$. The complete integration of (36) where $G = 1$ shows that by applying a proper similitude, the linear-element (35) of the sphere Σ may be written as

$$dS^2 = \operatorname{sech}^2 x\,(dx^2 + dy^2). \tag{37}$$

By applying the transformation $U = \log u^n$ to this, we find that the other important conformal map (17) of a sphere Σ upon a plane Π may be written as

$$dS^2 = \frac{4n^2(x^2 + y^2)^{n-1}(dx^2 + dy^2)}{[(x^2 + y^2)^n + 1]^2}. \tag{38}$$

The first map (37) represents a Mercator projection and the second map (38) is a stereographic (Ptolemy) projection for $n = 1$, and a Lambert conical projection for $n \neq 1$.

Thus if a sphere is mapped conformally upon a plane such that its scale curves form an isothermal family, then the conformal map must be either a Mercator, or a Ptolemy stereographic projection, or a Lambert projection. Also these are the only conformal maps of a sphere upon a plane such that the scale curves form a parallel family. A Mercator projection is the only conformal map of a sphere upon a plane such that its scale curves are straight lines. Finally stereographic projections and Lambert conical projection are the only conformal maps of a sphere upon a plane such that their scale curves are circles.

It is impossible to map a sphere conformally on a plane so that all the ∞^2 great circles become straight. But it can be shown from our work, that the three classic maps (Ptolemy, Mercator, Lambert) are the only conformalities which map the maximum possible number of geodesics (namely ∞^1) into straight lines.

Non-conformal maps of a general surface and of a sphere will be studied in our next paper. The scale curves then form a two-parameter family, so the theory is radically different. See *Science*, vol. 98 (1943), pp. 324-325.

COLUMBIA UNIVERSITY,

ILLINOIS INSTITUTE OF TECHNOLOGY.

GIBBS' PHENOMENON AND THE PRIME NUMBER THEOREM.*

By Aurel Wintner.

If $\epsilon > 0$ and $T > \epsilon$ are fixed, the series

$$(1) \qquad \sum_p p^{-1} \sin (t \log p),$$

being the imaginary part of

$$\sum_p p^{-1-it},$$

is uniformly convergent for $\epsilon < t < T$ (and therefore for $-T < t < -\epsilon$); in particular, (1) is convergent for every t. According to Mertens [2], these facts are substantially equivalent to the non-vanishing of $\zeta(s)$ on the line $\sigma = 1$, and therefore to the prime theorem (Ikehara). It is understood that p runs *increasingly* through all primes.

Since $\sum_p p^{-1} = \infty$, the series

$$(2) \qquad \sum_p p^{-1} \cos (t \log p)$$

cannot be uniformly convergent near $t = 0$ (it is uniformly convergent for $\epsilon < t < T$, since it is the real part of $\sum p^{-1-it}$). Due to the saltus of $\arg \zeta (1 + it)$ at $t = 0$, where $\zeta (s) \sim (s - 1)^{-1}$ and $s = 1 + it$, *the series* (1) *is not uniformly convergent near* $t = 0$.

This suggests a certain analogy between (2), (1) and the classical Fourier series

$$\sum_{n=1}^{\infty} n^{-1} \cos nt, \qquad \sum_{n=1}^{\infty} n^{-1} \sin nt.$$

In fact, the latter Fourier series belong to the Maclaurin series of the logarithm of $(1 - z)^{-1}$ in the same way as (2), (1) belong (except for trivial terms) to the Dirichlet series of the logarithm of $\zeta(s) \sim (s - 1)^{-1}$, where $z = e^{it}$ and $s = 1 + it$ respectively. Moreover, both series (2), (1) are Fourier series in the sense (B^2) and belong, in this sense, precisely to the functions to which they converge (if $t \neq 0$); cf. Wintner [4].

In view of this parallelism, one will expect that, corresponding to the classical Gibbs phenomenon of the series

---

* Received November 17, 1943.

$$\sum_{n=1}^{\infty} n^{-1} \sin nt,$$

the partial sums of the series (1) *are uniformly bounded near* $t = 0$.

More specifically, it will be shown that *the partial sums of the series* (1) *exhibit a Gibbs phenomenon at* $t = 0$; *in the sense that, as* $m \rightarrow \infty$, *the difference*

$$\sum_{p \leqq m} p^{-1} \sin (t \log p) - S(t \log m) \tag{3}$$

tends to a limit uniformly for $|t| < T$, *where* T *is arbitrarily fixed and*

$$S(x) = \int_0^x u^{-1} \sin u \, du \tag{4}$$

(so that $S(\pm \infty) = \pm \frac{1}{2}\pi$).

The proof is suggested by the formal remark that the ordinary Dirichlet series

$$-\sum_{n=2}^{\infty} (\log n)^{-1} n^{-s} \tag{5}$$

and the (aperiodic) trigonometric series (1) are substantially the integrals of

$$\sum_{n=1}^{\infty} n^{-s} = \zeta(s) \sim (s-1)^{-1}, \qquad (s \rightarrow 1), \tag{6}$$

and of the imaginary part of $\log \zeta(1+it)$, respectively. Correspondingly, (1) will first be replaced by

$$\sum_{n=2}^{\infty} (n \log n)^{-1} \sin (t \log n), \tag{7}$$

that is, by the imaginary part of (5) on the line $\sigma = 1$, where $s = \sigma + it$. It will then be shown that the uniformity statement made with regard to (3) remains unaffected if the partial sums of (1) occurring in (3) are replaced by the corresponding partial sums of (7).

It is a standard fact, readily verified by partial summation, that the series

$$\sum_{n=1}^{\infty} n^{-1-it}$$

of $\zeta(1+it)$ is divergent for every t and that the partial sums of this series are uniformly bounded for $\epsilon < t < T$, if $0 < \epsilon < T < \infty$ (but not if $\epsilon = 0$). Since $(\log n)^{-1}$ tends decreasingly to 0, it follows from Dedekind's convergence criterion (that is, again by partial summation) that the series (5), where $s = 1 + it$, and therefore also the imaginary part, (7), of this series, is uniformly convergent for $\epsilon < t < T$.

In order to control the partial sums of (7) as $\epsilon \rightarrow 0$, use will be made of an observation made in 1862 by Kinkelin, according to which

$$\zeta(1 + it) = \sum_{n=1}^{m} n^{-1-it} + (it\, m^{it})^{-1} + o(1) \quad \text{as} \quad m \rightarrow \infty$$

holds *uniformly* on every *fixed* open interval $\epsilon < t < T$, as seen by partial summation (deeper results, valid *within* the critical strip, are contained in the approximate functional equation of Riemann and Hardy-Littlewood; cf. [1]). Hence, if $\zeta^*(s)$ denotes the entire function

$$\zeta^*(s) = \zeta(s) - 1 - (s - 1)^{-1},$$

it follows, on writing u instead of t, that

$$\zeta^*(1 + iu) - iu^{-1} = \sum_{n=2}^{m} n^{-1-iu} - im^{-iu}u^{-1} + o(1) \quad \text{as} \quad m \rightarrow \infty$$

holds uniformly for $0 < u < T$, if T is fixed. For the real parts, this implies that

$$\Re \sum_{n=2}^{m} n^{-1-iu} = u^{-1} \sin(u \log m) + \Re\zeta^*(1 + iu) + o(1) \quad \text{as} \quad m \rightarrow \infty,$$

where the o-term is uniform for $0 < u < T$. Consequently, as $m \rightarrow \infty$,

$$\int_0^t \Re \sum_{n=2}^{m} n^{-1-iu} du - \int_0^t u^{-1} \sin(u \log m)\, du \rightarrow \int_0^t \Re\zeta^*(1 + iu)\, du$$

holds uniformly for $0 < u < T$. But the second integral on the left is $S(t \log m)$, by (4); while the first integral on the left is the m-th partial sum of the series (7), since, in accordance with the remark made in connection with (5),

$$-i \int_0^t \sum_{n=2}^{m} n^{-1-iu}\, du = \sum_{n=2}^{m} (n \log n)^{-1}(n^{-it} - 1).$$

Accordingly, as $m \rightarrow \infty$,

$$(8) \qquad \sum_{n=2}^{m} (n \log n)^{-1} \sin(t \log n) - S(t \log m) \rightarrow \int_0^t \Re\zeta^*(1 + iu)\, du$$

holds uniformly for $0 < t < T$, where T is arbitrarily fixed.

It follows that, in order to prove that the expression (3) tends to a limit uniformly for $0 < t < T$, it is sufficient to show that, as $m \rightarrow \infty$, the difference between the first term, Σ, of the expression (3) and the m-th partial sum of the series (7) tends to 0 uniformly for $0 < t < T$. In other words, it is sufficient to prove that the series

$$(9) \qquad \sum_{k=1}^{\infty} a_k \sin(t \log k)$$

is uniformly convergent for $0 < t < T$, where the coefficients a_k are defined by

$$(10) \qquad \sum_{k=1}^{\infty} a_k k^{-s} = \sum_p p^{-s} - \sum_{n=2}^{\infty} (\log n)^{-1} n^{-s}, \qquad (\sigma > 1).$$

But if $s = 1 + it$, then (9) is the negative of the imaginary part of (10). Consequently, it is sufficient to show that the Dirichlet series (10) is uniformly convergent on every fixed segment $|t| < T$ of the line $\sigma = 1$. And this can be proved as follows:

Since $\zeta(s) = \prod_p (1 - p^{-s})^{-1}$ for $\sigma > 1$,

$$\sum_p p^{-s} = \log \zeta(s) + g(s),$$

where $g(s)$ is a function regular for $\sigma > \frac{1}{2}$. On the other hand, since $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ for $\sigma > 1$,

$$\sum_{n=2}^{\infty} (\log n)^{-1} n^{-s} = \int_2^s \zeta(z)\, dz - s + \text{const.}$$

for $\sigma > 1$. Hence, from (10),

$$(11) \qquad \sum_{k=1}^{\infty} a_k k^{-s} = \log \zeta(s) + \int_2^s \zeta(z)\, dz + f(s), \qquad (\sigma > 1),$$

where $f(s)$ is a function regular for $\sigma > \frac{1}{2}$. But

$$(12) \qquad \zeta(s) = (s-1)^{-1} + \text{an entire function}$$

and, by the function-theoretical formulation of the prime number theorem,

$$(13) \qquad \zeta(s) \neq 0 \quad \text{for} \quad \sigma \geqq 1.$$

Since (12) and (13) obviously imply that the sum of the first two terms on the right of (11) is regular for $\sigma \geqq 1$, and since the third term, $f(s)$, is regular for $\sigma > \frac{1}{2}$ and therefore for $\sigma \geqq 1$, it follows from (11) that the difference on the right of (10) is regular for $\sigma \geqq 1$.

This is the relevant consequence of the prime number theorem. In fact, the balance of the proof will in no sense involve the prime number theorem.

According to the analogue of the Tauberian theorem of P. Fatou–M. Riesz for Dirichlet series

$$\sum_{k=1}^{\infty} a_k k^{-s},$$

(cf. M. Riesz [3]), such a series must converge uniformly on every fixed segment $-T < t < T$ of the line $\sigma = 1$ if, on the one hand,

$$\sum_{k=1}^{x} a_k = o(x) \quad \text{as} \quad x \to \infty \tag{14}$$

and, on the other hand, the Dirichlet series converges in the open half-plane $\sigma > 1$ to a function which remains regular on the line $\sigma = 1$. Since the latter assumption has just been assured for (10), all that remains to be proved is that (14) is satisfied when a_k is defined by (10).

It is clear from (10) that the sum (14) is now identical with

$$\sum_{k=1}^{x} a_k = \sum_{p \leqq x} 1 - \sum_{n=2}^{x} (\log n)^{-1}. \tag{15}$$

But the first term on the right of (15) is the number of primes not exceeding x, and therefore, according to Chebyshev, it is majorized by a constant multiple of $x/\log x$, which is $o(x)$. Since the second term on the right of (15) is

$$\sum_{n=2}^{x} (\log n)^{-1} = \sum_{n=2}^{x} o(1) = o(x),$$

the proof is complete.

Since only the roughest estimates were needed in this verification of (14), one might have the impression that the application of a Tauberian theorem can be avoided, if the formulation $\pi(x) \sim x/\log x$ of the prime number theorem is used in (15). In fact, if (15) is written in the form

$$\sum_{k=1}^{x} a_k \doteq \pi(x) - \sum_{n=2}^{x} (\log n)^{-1} \sim \pi(x) - \int_{2}^{x} (\log u)^{-1} du \sim \pi(x) - x/\log x,$$

then $\pi(x) \sim x/\log x$ improves the $o(x)$ in (14) to $o(x/\log x)$. However, this does not suffice for the elimination of the Tauberian theorem. What would suffice is the improvement of the estimate $o(x/\log x)$ to $o(x/\log^{1+\epsilon} x)$ for some $\epsilon > 0$, since the proof could then be based on

$$\int_2^\infty |\, o(x/\log^{1+\epsilon} x) \,|\, dx < \infty, \tag{17}$$

where $\epsilon > 0$ could be arbitrarily small. But the fact is that, while even de la Vallée–Poussin's zero-free domain implies much more than the truth of $\pi(x) - x/\log x = o(x/\log^\beta x)$ for some $\beta > 1$, the prime number theorem itself, which is precisely (13), supplies only $\beta = 1$, and therefore no $\beta = 1 + \epsilon > 1$ satisfying (17).

This methodical situation is pointed out because the above proof is such as to avoid the prime number theorem entirely, if only the question of Gibbs' phenomenon, i. e., the vicinity of the point $t = 0$, is concerned. In fact, it then suffices to apply the Tauberian theorem for a *fixed* (small) T.

On the other hand, if T is allowed to be arbitrarily large, then the above procedure is reversible, i. e., *the uniform convergence of the Dirichlet series* (10) *on every fixed segment* $-T < t < T$ *of the line* $\sigma = 1$ *is equivalent to the prime number theorem,*

$$p_n \sim n \log n. \tag{18}$$

In fact, if the Dirichlet series (10) converges on the line $\sigma = 1$ to a continuous function, then it is clear from (11) that (12) implies (13).

In this connection, it is worth mentioning that *Riemann's hypothesis is equivalent to the convergence of the Dirichlet series* (10) *for* $\sigma > \frac{1}{2}$. In fact, Riemann's hypothesis is equivalent to the truth of

$$\pi(x) - \mathrm{Li}\,(x) = o(x^{\frac{1}{2}+\epsilon}), \quad \text{where} \quad \mathrm{Li}\,(x) \equiv \int_2^x (\log u)^{-1}\, du, \tag{19}$$

for every $\epsilon > 0$. But (19) is just a rewording of the statement that the abscissa of convergence of the Dirichlet series (10) does not exceed $\frac{1}{2} + \epsilon$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] G. H. Hardy and J. E. Littlewood, "The zeros of Riemann's zeta-function on the critical line," *Mathematische Zeitschrift*, vol. 10 (1921), pp. 283-317.

[2] F. Mertens, "Ueber die Konvergenz einer aus Primzahlpotenzen gebildeten unendlichen Reihe," *Göttinger Nachrichten* (1887), pp. 265-269.

[3] M. Riesz, "Ein Konvergensatz für Dirichletsche Reihen," *Acta Mathematica*, vol. 40 (1916), pp. 349-361.

[4] A. Wintner, "The behavior of Euler's product on the boundary of convergence," *Duke Mathematical Journal*, vol. 10 (1943), pp. 429-440.

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

NUMBER-THEORETICAL ALMOST-PERIODICITIES.*

By Aurel Wintner.

The present paper deals with an harmonic analysis of various types of "hidden periodicities" in the theory of whole numbers. The "periodicities" in question, the simplest manifestations of which have so much fascinated the Greeks, are not of course periodicities proper, and so it is remarkable that the corresponding arithmetical recurrences possess at all an harmonic analysis in the technical sense of the term. In fact, the "recurrences" prove to correspond to an almost-periodic behavior in the sense (B) of Besicovich. It is revealing for the whole situation that, for instance, Bohr's class of uniformly almost-periodic sequences turns out to be all too restrictive to represent the effect of arithmetical fluctuations, except in trivial cases (cases which cannot even occur in connection with the "Eulerian" and "Dirichletian" algorithms, treated in **5** and **9** respectively).

The emphasis in the results obtained is always on the circumstance that the resulting expansions into *trigonometric series,* instead of claiming formal identities, represent *Fourier expansions* for the arithmetical sequences to which they belong. Otherwise the results would have nothing to do with an *harmonic analysis* of the "hidden periodicities" in arithmetics.

1. By a function h will be meant a function $h = h(n)$ or $h = h_m$ of a positive integer, that is, a sequence $h(1), h(2), \cdots$ or $h_1, h_2, \cdots$. The symbol $M(h)$ will denote the mean-value

$$M(h) = \lim_{n\to\infty} (1/n) \sum_{m=1}^{n} h(m), \tag{1}$$

provided that (1) exists as a *finite* limit.

Consider a linear transformation

$$f(n) = \sum_{m=1}^{\infty} e_m(n) g(m) \tag{2}$$

of an arbitrary function g into a function f. The latter is defined for an arbitrary g if and only if the matrix of (2) is "zeilenfinit" (Toeplitz), that is to say such that there exists a function $N = N_m$ having the property that

$$e_m(n) = 0 \quad \text{whenever} \quad n > N_m, \qquad (m = 1, 2, \cdots). \tag{3}$$

---

* Received February 14, 1944.

For instance, (3) is satisfied by $N_m = m$ if the matrix of (2) is "triangular".

Suppose that the matrix of (2) satisfies (3) and is such that its elements contained in any fixed column form an almost-periodic function (B) of the index of the row, that is, that the functions

$$(4) \qquad e_1(n), e_2(n), \cdots, e_m(n), \cdots \text{ are almost-periodic } (B)$$

(the variable is n, while m is arbitrarily fixed). Then the absolute value of each of the functions (4) of n is almost-periodic (B). In particular, the mean-value $M(|e_m|)$, defined by the case $h = e_m$ of (1), exists for $m = 1, 2, \cdots$.

If the matrix of the linear transformation (2) satisfies (3) and (4), it is easy to verify that any function g satisfying

$$(5) \qquad \sum_{m=1}^{\infty} M(|e_m|) |g(m)| < \infty$$

is transformed by (2) into a function $f(n)$ which is almost-periodic (B) and such that every Fourier constant of $f(n)$ is the limit, as $k \rightarrow \infty$, of the corresponding Fourier constant of $f_k(n)$, where

$$(6) \qquad f_k(n) = \sum_{m=1}^{k} e_m(n) g(m);$$

in fact, f_k will tend, as $k \rightarrow \infty$, to f in the mean of the (B)-space (cf. [11], pp. 24-25).

Incidentally, it is clear from the proof of these facts that the full force of the assumption (3), an assumption which is chosen independent of g, is not needed.

2. Suppose that the Fourier series (B) of each of the functions (4) is periodic and that the m-th of these Fourier series (B) has the period m, where $m = 1, 2, \cdots$. This means that there exist m constants $\alpha_m(1), \cdots, \alpha_m(m)$ satisfying

$$(7) \qquad e_m(n) \sim \sum_{j=1}^{m} \alpha_m(j) \epsilon(nj/m),$$

where m is arbitrarily fixed and

$$(8) \qquad \epsilon(x) = \exp(2\pi i x).$$

On the other hand, none of the functions (4) of n is required to be periodic. In other words, (7) is compatible with

$$(9) \qquad e_m(n) \neq e^m(n),$$

if $e^m(n)$ denotes the sum on the right of (7). All that can be said is that, besides (7),

$$e^m(n) \sim \sum_{j=1}^{m} \alpha_m(j)\epsilon(nj/m), \tag{10}$$

since

$$e^m(n) = \sum_{j=1}^{m} \alpha_m(j)\epsilon(nj/m) \tag{11}$$

is a finite, and therefore a uniformly convergent, trigonometric sum in n. Correspondingly, the Fourier constants $\alpha_m(j)$ can be calculated not only by the Fourier inversion

$$m\alpha_m(j) = \sum_{n=1}^{m} e_m(n)\epsilon(-jn/m) \tag{12}$$

of (7) but also by the Fourier inversion

$$m\alpha_m(j) = \sum_{n=1}^{m} e^m(n)\epsilon(-jn/m) \tag{13}$$

of (11).

For the purposes of **5** below, it will be fundamental that the possibility (9) is not excluded by the assumptions of the following theorem:

(I) *If the matrix of a linear transformation* (2) *satisfies* (3) *and the particular case* (7) *of* (4), *and if g is any function satisfying* (5), *then the function f into which* (2) *transforms g is almost-periodic* (*B*) *and has a limit-periodic Fourier series* (*B*),

$$f(n) \sim \sum_{r=1}^{\infty} \sum_{1 \leqq s \leqq r}^{(s,r)=1} a_r{}^s \epsilon(ns/r), \tag{14}$$

in which the Fourier constants have the values represented by the series

$$a_r{}^s = \sum_{r|k} g(k)\alpha_k(ks/r) \tag{15}$$

(*each of which is absolutely convergent*); *so that, in particular,*

$$M(f) = a_1{}^1 = \sum_{k=1}^{\infty} g(k)\alpha_k(k). \tag{16}$$

It is clear from (7) and from the definition of a Fourier constant that $|\alpha_m(j)| \leqq M(|e_m|)$ for every j. On the other hand, the summation index k in (5) runs through all multiples of r, that is, the series (15) can be written in the form

$$\sum_{n=1}^{\infty} \alpha_{nr}(ns)g(nr).$$

Hence, if the preceding inequality is applied to $m = nr$ and $j = ns$, the series (15) is seen to be majorized by the series

$$\sum_{n=1}^{\infty} M(|e_{nr}|)|g(nr)|.$$

But the latter series is a subseries of the series (5), since r is any fixed positive integer. This proves the parenthetical assertion following (15).

Of the remaining statements of (I), only the explicit representation (14), (15) of the Fourier analysis (B) of f remains to be proved. In fact, the almost-periodicity (B) of f follows from **1**, even if the restriction (10) of (4) is omitted. Correspondingly, the proof of (I) will be complete if it is shown that, in virtue of the restriction (10), the function (6) has a Fourier series (B) of the form

$$(17) \qquad f_k(n) \sim \sum_{r=1}^{\infty} \sum_{1 \leqq s \leqq r}^{(s,r)=1} a_r{}^s(k) \epsilon(ns/r)$$

(for every fixed k) and that, as $k \to \infty$, each of the Fourier constants of (17) tends to the limit

$$(18) \qquad \lim_{k \to \infty} a_r{}^s(k) = \sum_{n=1}^{\infty} \alpha_{nr}(ns) g(nr),$$

that is, to the value (15). In fact, the remark following (6) implies that (17) and (18) are sufficient for (14) and (15).

Substitution of (10) into (6) gives

$$(19) \qquad f_k(n) \sim \sum_{m=1}^{k} \sum_{j=1}^{m} \alpha_m(j) \epsilon(nj/m).$$

This can be written in the form (17), if all those pairs j, m are collected in (19) for which the quotient j/m has a fixed value s/r, where s and r are relatively prime. In other words, (17) is satisfied by

$$(20) \qquad a_r{}^s(k) = \sum_{j/m = s/r}^{m \leqq k} \alpha_m(j).$$

But every pair of summation indices j, m occurring in (20) can be represented in the form $j = ns$, $m = nr$, where n is a *unique* positive integer. It is also clear that this n will attain *all* values $1, 2, \cdots$ as $k \to \infty$ in (20). Hence, (18) follows from (20) and from the fact that, as shown above, the series on the right of (18) is *absolutely* convergent.

This completes the proof of (I).

3. Suppose that the matrix of (2) satisfies the assumptions, (2), (4) and (7), of (I).

Of particular arithmetical interest proves to be the case of those among the matrices which have the property that, for any function g satisfying (5), the $\phi(r)$ coefficients (15) of each of the interior sums in (14) are identical except for *fixed phase factors.* By this is meant that there should exist for every s and every r, where

(21) $$(s, r) = 1 \quad \text{and} \quad 1 \leqq s \leqq r, \qquad (r = 1, 2, \cdots),$$

a constant $\alpha_r{}^s$ which depends only on the matrix of (2) and has the property that, on the whole s-range (21) belonging to any fixed r, the Fourier constants (15) are expressible in the form

(22) $$a_r{}^s = \epsilon(-\alpha_r{}^s)b_r,$$

where the coefficients $b_1, b_2, \cdots$ are independent of s, and g in (15) is any function satisfying (5). In other words, the Fourier series (14) should have the particular structure

(23) $$f(n) \sim \sum_{r=1}^{\infty} b_r \beta_r(n),$$

where the trigonometric sums

(24) $$\beta_r(n) = \sum_{1 \leqq s \leqq r}^{(s,r)=1} \epsilon(-\alpha_r{}^s + ns/r)$$

depend only on the matrix of (2).

According to (22) and (15), the matrix of (2) will have this property if and only if there exists for every function $g = g(m)$ satisfying (5) a function $b = b_m$ satisfying

(25) $$b_r = \epsilon(\alpha_r{}^s) \sum_{r|k} g(k)\alpha_k(ks/r),$$

where, for any fixed r, the integer s is restricted only by (21). Clearly, (25) can be written in the form.

(26) $$0 = \sum_{n=1}^{\infty} g(rn)\{\epsilon(\alpha_r{}^s)\alpha_{nr}(ns) - \epsilon(\alpha_r{}^t)\alpha_{nr}(nt)\},$$

where t denotes any fixed s satisfying (21) with reference to a given r. But (26) is required for every g satisfying (5) and, therefore, for every $g = g(m)$ tending sufficiently fast to 0 as $m \to \infty$. Since the expression $\{\quad\}$ multiplying $g(rn)$ in (26) depends only on the matrix of (2), it follows that every $\{\quad\}$ in (26) must vanish. But t is a fixed s satisfying (21). Hence, the vanishing of every $\{\quad\}$ in (26) means that, if r, n is any pair of positive integers, the product $\epsilon(\alpha_r{}^s)\alpha_{nr}(ns)$ has a value independent of the choice of the $\phi(r)$ integers s satisfying (21). If r and nr are denoted by d and m, respectively, then this necessary and sufficient condition for the phase rule (22) appears in the following form:

For every positive integer, m, and for every divisor, d, of m, there exists a value, $\lambda = \lambda_m(d)$, such that

(27) $$\epsilon(\alpha_d{}^s)\alpha_m(sm/d) = \lambda_m(d)$$

holds for all $\phi(d)$ integers satisfying

$$(s,d)=1 \quad \text{and} \quad 1 \leqq s \leqq d, \quad \text{where} \quad d \mid m. \tag{28}$$

This criterion can readily be applied to an explicit determination of the linear transformations (2) in question. In fact, the number of pairs of integers s, d, satisfying (28) for a fixed m is

$$\sum_{d|m} \sum_{1 \leqq s \leqq d}^{(s,d)=1} 1 = \sum_{d|m} \phi(d) = m.$$

Correspondingly, the Fourier expansion (7) of m terms can be written in the form

$$e_m(n) \sim \sum_{d|m} \sum_{1 \leqq s \leqq d}^{(s,d)=1} \alpha_m(sm/d)\epsilon(ns/d), \tag{29}$$

whether (27) be satisfied or not. Consequently, the Fourier constants $\alpha_m(j)$ of the columns (4) of the matrix of (2) satisfy (27) if and only if the Fourier expansions (7) can be represented in the form

$$e_m(n) \sim \sum_{d|m} \lambda_m(d) \sum_{1 \leqq s \leqq d}^{(s,d)=1} \epsilon(-\alpha_d{}^s + ns/d),$$

which means, by (24), that

$$e_m(n) \sim \sum_{d|m} \lambda_m(d)\beta_d(n). \tag{30}$$

Since (27) reduces (25) to

$$b_r = \sum_{r|k} \lambda_k(r) g(k), \tag{31}$$

the result of these considerations may be summarized as follows:

(II) *Let the matrix of a linear transformation* (2) *satisfy the assumptions,* (3), (4) *and* (7), *of* (I). *Then the Fourier series* (14) *will have the particular form* (23), (24) *belonging to a given set of* (*not necessarily real*) *absolute constants* $\alpha_r{}^s$ *if and only if there exist absolute constants* $\lambda_m(n)$ *by means of which the Fourier series of the columns* (4) *of the matrix are expressible in the form* (30); *in which case the coefficients of the Fourier series* (23) *of the almost-periodic* (*B*) *functions* $f(n)$ *are given by the absolutely convergent series* (31), *where* $g(n)$ *denotes the function, subject to* (5), *which defines* $f(n)$ *by* (2).

4. Perhaps the simplest particular case results if (2) is the linear transformation corresponding to the sieve of Eratosthenes, namely,

$$f(n) = \sum_{d|n} f'(d), \tag{32}$$

where

$$g(n) = f'(n). \tag{33}$$

The functional notation (33) in the case (32) of (2) is justified by the fact that the function (33) is uniquely determined by (32), if $f(n)$ is any given function. In fact, if (32) is written in the form (2), then (3) is satisfied by $N_m = m$, and the diagonal element $e_m(m)$ is 1 for every m. Thus, the linear transformation (32) of f' into f has a unique inverse, which proves the assertion.

It is also seen that, if (32) is identified with (2) in virtue of (33), then the m-th of the functions (4) can be described as follows: For every m, the function $e_m(n)$ of n has the period m and is 0 or 1 according as $1 \leqq n < m$ or $n = m$ (in fact, this description of the matrix of (32) is equivalent to the sieve process of Eratosthenes). But (8) shows that this description of the functions (4) is equivalent to the explicit formula

$$e_m(n) = (1/m) \sum_{j=1}^{n} \epsilon(nj/m). \tag{34}$$

According to (7) and (11), this means that

$$\alpha_m(j) = 1/m \tag{35}$$

and

$$e_m(n) = e^m(n). \tag{36}$$

Finally, it is seen from (1), (7) and from the case $j = m$ of (35), that $M(e_m) = 1/m$.

Thus it is clear that, if (2) is given by (32) and (33), then the assertion of (I) can be formulated as follows:

(III) *There exists for every function f a unique function f' by means of which f is representable in the form* (32). *Suppose that the function f' belonging to f satisfies the condition*

$$\sum_{k=1}^{\infty} |f'(k)|/k < \infty. \tag{37}$$

Then the function f is almost-periodic (B) *and has the Fourier series* (B)

$$f(n) \sim \sum_{r=1}^{\infty} a_r c_r(n), \tag{38}$$

where $c_r(n)$ denotes the Ramanujan sum

$$c_r(n) = \sum_{1 \leqq s \leqq r}^{(s,r)=1} \epsilon(ns/r) \tag{39}$$

and the Fourier constants a_r *are given by the absolutely convergent series*

$$a_r = \sum_{r|k} f'(k)/k; \tag{40}$$

so that, in particular,

$$M(f) = a_1 = \sum_{k=1}^{\infty} f'(k)/k. \tag{41}$$

This particular case of (I) was found in [11], pp. 24-25 and p. 30. Considerations with regard to an "arbitrary" function f which sound like (III) but actually either deal with trigonometric identities, rather than with Fourier series, or postulate, rather than prove, the existence of a Fourier series (38) of some sort are, of course, of a much older date; cf., e. g., the considerations of Libri [8] or Carmichael [2]. Incidentally, it is sufficient to compare (14) in (I) or (23) in (II) with the uniqueness theorem of a Fourier series (B), in order to see how meaningless a postulation of the *specific* expansion (38) in (III) would be. In this connection, cf. a question raised by Hardy [6].

5. Let now the "Eratosthenian" process (32) be replaced by the "Eulerian" process

$$f(n) = \sum_{1 \leqq l \leqq n}^{(l,n)=1} f^*(l) \tag{42}$$

("Eulerian," since, in virtue of (42),

$$f(n) = \phi(n) \quad \text{when} \quad f^*(n) = 1; \tag{43}$$

incidentally, from (39),

$$f(n) = c_r(n) \quad \text{when} \quad f^*(n) = \epsilon(nr), \tag{44}$$

where r is any fixed positive integer).

It is clear from (42) that $f(1) = f^*(1)$ and $f(2) = f^*(1)$; hence

$$f(1) = f(2). \tag{45}$$

Conversely, if $f = f(n)$ is any function satisfying the initial restriction (45), then there exists a unique function $f^* = f^*(n)$ by means of which f is representable in the form (42). In other words, the functional notation

$$g = f^* \tag{46}$$

is justified if (42) is thought of as written in the form (2), provided that the trivial necessary condition (45) is satisfied. In fact, if $n \geqq 2$, then (l,n) is 1 for $l = n - 1$ but is not 1 for $l = n$. It follows, therefore, from (42) that, if $n \geqq 2$, then $f(n) = F_{n-2} + f^*(n-1)$, where F_{n-2} denotes the sum of certain of the $n-2$ values $f^*(m)$ belonging to $m = 1, \cdots, n-2$ (with the understanding that the number of these values can be 0, in which case $F_{n-2} = 0$). Consequently, if (45) is satisfied, then, since $f^*(n-1) = f(n) - F_{n-2}$ for $n \geqq 2$ and $f^*(1) = f(1)$, $f^*(2) = f(1)$, the values of f^* can be determined, recursively, from those of f for every n. This proves the assertion.

Let (42) be thought of as written in the form (2), where the function (46) is arbitrary. Then, since F_{n-2} does not contain $f(m)$ when $m > n-1 (> 0)$, condition (3) is satisfied by $N_m = m-1$ if $n \geqq 2$, while $N_1 = 1$. Actually, the sieve process of Eratosthenes, referred to in connection with (32), can readily be adapted to (42) and leads, when (42) is written in the form (2), to the following description of the structure of the column functions (4): For every positive integer m (including $m = 1$),

$$(47) \quad e_m(n) = 0 \quad \text{if} \quad 1 \leqq n < m \quad \text{and} \quad e_m(n) = e^m(n) \quad \text{if} \quad m \leqq n < \infty,$$

where $e^m(n)$ denotes 1 or 0 according as n and m are or are not relatively prime (so that, in particular,

$$(48) \qquad e_1(n) = e^1(n) = 1$$

for every n).

Clearly, the function $e^m(n)$ just defined is a periodic function of n and has m as a period. Although the primitive period of $e^m(n)$ is in general (namely, unless m is square-free) less than m, it follows that $e^m(n)$ can, for every m, be represented in the form (11). The coefficients $\alpha_m(j)$ of (11) can be calculated from (13). This gives

$$m\alpha_m(j) = \sum_{1 \leqq s \leqq m}^{(s,m)=1} \epsilon(-js/m).$$

But the sum on the right of this relation remains unaltered if $-j$ is replaced by j. Hence, from (39),

$$(50) \qquad \alpha_m(j) = c_m(j)/m.$$

It follows, therefore, from (11) that the function $e^m(n)$ occurring in the representation (47) of the matrix of the linear transformation (42) of f^* into f can be given explicitly as follows:

$$(51) \qquad e^m(n) = \sum_{j=1}^{m} c_m(j)\epsilon(nj/m)/m.$$

For the column functions (4) themselves, (51) implies the Fourier series (B)

$$(52) \qquad e_m(n) \sim \sum_{j=1}^{m} c_m(j)\epsilon(nj/m)/m,$$

by (11), (10) and (7); cf. (9) and (47). In particular,

$$M(|e_m|) = M(e_m) = \phi(m)/m. \tag{53}$$

In fact, the constant term of the Fourier series (52) belongs to $j = m$ and has, therefore, the value $c_m(m)/m$. This means, by (1), that $M(e_m) = c_m(m)/m$. Hence, (53) follows by placing $r = m$ and $n = m$ in (39) and using the fact that $e_m \geqq 0$.

If all of this is combined with (I), there results the following theorem:

(IV) *For an arbitrary function $f = f(n)$ subject to* (45), *there exists a unique function $f^* = f^*(n)$ by means of which f is representable in the form* (42). *Suppose that the function f^* belonging to f satisfies the condition*

$$\sum_{k=1}^{\infty} |f^*(k)|\,\phi(k)/k < \infty, \tag{54}$$

where ϕ denotes Euler's function. Then f is almost-periodic (B) *and has the Fourier series* (B)

$$f(n) \sim \sum_{r=1}^{\infty} a_r c_r(n), \tag{55}$$

where the $c_r(n)$ are the Ramanujan sums (39) *and, if μ denotes Möbius' function, the Fourier constants a_r are given by the absolutely convergent series*

$$a_r = \frac{\mu(r)}{\phi(r)} \sum_{r|k} \frac{\phi(k)}{k} f^*(k); \tag{56}$$

so that, in particular,

$$M(f) = a_1 = \sum_{k=1}^{\infty} f^*(k)\phi(k)/k \tag{57}$$

and

$$a_r = 0 \text{ unless } r \text{ is square-free.} \tag{58}$$

In fact, if (2) is identified with (42), it is seen from (46) and (53) that the assumption (5) of (I) becomes the assumption (54) of (IV). On the other hand, (46) and (50) show that the assertion (15) of (I) for (14) can now be written in the form

$$a_r{}^s = \sum_{k=1}^{\infty} f^*(rk) c_{rk}(ks)/(rk), \tag{59}$$

where $(s, r) = 1$, by (21) or (14). But it is easily verified from (39) and from the definitions of Euler's ϕ and of Möbius' μ that

$$c_{rk}(ks) = \phi(kr)\mu(r)/\phi(r), \quad \text{where} \quad (s, r) = 1, \tag{60}$$

holds for $k = 1, 2, \cdots$; cf. (61) below. Consequently, (59) can be written in the form $a_r{}^s = a_r$, if a_r is defined by (56). Finally, since $a_r{}^s = a_r$ holds for all $\phi(r)$ integers s satisfying (21), it is seen from (39) that (14) becomes precisely (55).

This completes the proof of (IV). The crucial point was, of course, that, just as in (III), the $\phi(r)$ coefficients of the interior sum in (14) proved to be independent of s. Correspondingly, both (III) and (IV) can be thought of as paradigmata of (II), the *class functions* $\beta_r(n)$ of the *representations* compatible with (5) being the functions $c_r(n)$, $|\mu(r)|\, c_r(n)$ in the respective cases (III), (IV). In particular, $\beta_r(n)$ is missing in the second (but not in the first) case unless $|\mu(r)| = 1$, that is, unless r is square-free.

It may be mentioned that the elementary identity (60), used in the reduction of (59) to (56), is just the particular case of the relation

$$(61) \qquad c_m(n)/\phi(m) = \mu(m/(m,n))/\phi(m/(m,n)),$$

pointed out by Tegnér [10] for any pair of positive integers m, n and their least common divisor (m,n); cf. also Hölder [7].

6. Any function f satisfying (45) determines a unique pair of functions f', f^* by means of which f is representable in the respective forms (32), (42). Hence, (III) and (IV) supply two sufficient criteria for one and the same property, namely, the almost-periodicity (B), of an arbitrary f. In fact, the initial restriction (45), which is necessary in (IV), is quite immaterial, since, if $f(n)$ is given for every $n > 1$, the value of $f(1)$ can be chosen arbitrarily, a finite number of values $f(n)$ having no influence on the almost-periodicity (B) or on the Fourier expansion (B) of f.

It will now be shown that neither of the theorems (III), (IV) is contained in the other:

(V) *The two sufficient criteria which* (III) *and* (IV) *supply for the almost-periodicity* (B) *of a function f are independent, that is, condition* (37) *neither implies, nor is implied by, condition* (54).

In fact, if the values $f'(1), f'(2), \cdots$ are so chosen that (37) is satisfied and the numerical series (40) belonging to $r = 4$ has a sum distinct from 0, then, since (54) implies (58), it is clear that (37) cannot suffice for (54).

The remaining negation of (V) follows by choosing $f^*(n)$ to be 1 or 0 according as n is or is not 2. Then (54) is satisfied, since only one term of the series (54) is distinct from 0. It remains to show that this example does not satisfy (37).

First, since $f^*(2) = 1$ and $f^*(n) = 0$ for $n \neq 2$, it is seen from (42) that $f(n)$ is 1 or 0 according as n is or is not an odd number distinct from 1. Since (32) is equivalent to Möbius' inversion

$$f'(n) = \sum_{d|n} \mu(n/d) f(d) \tag{62}$$

for arbitrary mates f, f', it follows that the present $f'(n)$ is identical with the sum $\Sigma_1 \mu(n/d)$ in which d runs through all those odd divisors of n which are distinct from 1. Hence, if n is an odd prime p, it is clear that $f'(p)$ is represented by the single term $\mu(p/d) = \mu(p/p)$; so that $f'(p) = \mu(1) = 1$. Since the sum of the reciprocal values of all primes p is divergent, it follows that (37) is violated.

This completes the proof of (V).

Due to the representation $\Sigma \mu(n/d)$ of $f'(n)$ in the last example, it is possible to conclude from the prime number theorem that, in this example, the series

$$\sum_{k=1}^{\infty} f'(k)/k \tag{63}$$

is convergent (although (37) is not satisfied). This suggests the question as to whether or not (5) *always* implies the convergence (though not, of course, the absolute convergence) of (63). The answer proves to be affirmative. In view of (I) and (IV), much more than this assertion is contained in the following theorem:

(VI) *If the function $f(n)$ is almost-periodic (B), then* (63) *is a convergent series.*

The point in (VI) is that the assumption of the mere existence of $M(f)$, an assumption weaker than the almost-periodicity (B) of f, is insufficient for the convergence of (63). This was shown by an appropriate construction ([11], p. 13), while the question as to the sufficiency of the almost-periodicity (B) of f remained there undecided (cf. [11], p. 26).

If f is almost-periodic (B), the same is true of $|f|$. Hence, the mean-value (1) exists for both functions $h = f$, $h = |f|$. This means that $M(f_1)$ and $M(f_2)$ exist for the linear combinations $f_1 = |f| + f$, $f_2 = |f| - f$ of $f, |f|$. On the other hand, it is clear from the distributive character of the connection (32) or (62) between the f and f' that, if (63) converges when f is replaced by f_i, where $i = 1, 2$, then (63) must converge for $f = \frac{1}{2}(f_1 - f_2)$ itself. But $M(f_i)$ exists and $f_i \geqq 0$ holds for $i = 1, 2$, if, without loss of generality, f is assumed to be real-valued. Hence, in order to prove (VI), it is sufficient to ascertain that the existence of $M(f)$ and the assumption $|f| \geqq 0$ imply the convergence of (63). But the truth of this implication was recently established ([12], p. 6); it depends on somewhat more than the prime number theorem (cf. [12], p. 9).

Incidentally, (VI) itself contains the prime number theorem. In fact,

$f(n)$ is 1 or 0 according as $n=1$ or $n>1$, then $f(n)$ clearly is almost-periodic (B), with $0+0+\cdots$ as Fourier series, while (62) shows that $f'(n)$ becomes identical with $\mu(n)$. It follows therefore from (VI) that the series (63) converges when f' is Möbius' function μ. But this is known to be equivalent to the prime number theorem.

In the above proof of (VI), only the existence of $M(f)$ and $M(|f|)$ was needed. Hence, (VI) can be generalized as follows:

(VII) *If f is a function for which both $M(f)$ and $M(|f|)$ exist, then* (63) *is a convergent series.*

That (VII) *actually* is more general than (VI), follows from the existence of functions f which attain only the values 0 and 1 but fail to be almost-periodic (B), although $M(f)$ exists ([11], p. 28).

Remark. Corresponding to Möbius' inversion (62) of (32), there is an explicit inversion formula of (42), (45), namely,

$$(64)\quad f^*(n)=\sum_{d|n}f(d)-\sum_{d|n-1}f(d),\quad \text{where}\quad n>1;\qquad f^*(1)=f(1).$$

In fact, it is clear from (42) that

$$(65)\qquad \sum_{d|n}f(d)=\sum_{k=1}^{n}f^*(k).$$

But (65) is equivalent to (64).

Incidentally, (32) shows that (65) can be written in the form

$$(66)\qquad f(n)=F'(n),$$

where

$$(67)\qquad F(n)=\sum_{k=1}^{n}f^*(k).$$

Since (62) is equivalent to (32), the explicit form of (66) is

$$(68)\qquad f(n)=\sum_{d|n}\mu(n/d)F(d).$$

Thus (42) is identical with (68) in virtue of (67).

7. In view of (III) and (IV), it seems to be worthwhile to mention an instance of a Fourier expansion (14) which fails to be of the particular form of the Ramanujan series (38) occurring in (III), (IV), although it belongs to an almost-periodic (B) function $f(n)$ of classical arithmetical significance:

(VIII) *Let $f(n)$ denote the characteristic function of the set of the* *positive integers which are sums of three squares. Then* $f(n)$ *is almost-periodic* (B) *and has the limit-periodic Fourier series* (B)

$$(69) \qquad f(n) \sim 6^{-1} + \sum_{r=1}^{\infty} 2^{-r} \sum_{1 \leqq q \leqq 2^r} \lambda_q \epsilon(nq/2^r),$$

where q *is odd,* $\epsilon(x) = \exp(2\pi i x)$ *and*

$$(70) \qquad \lambda_q = -\log\{1 - \epsilon(q/8)\},$$

if the imaginary part of the logarithm is chosen between $\pm\,\pi$.

It is understood that by the characteristic function of a set of positive integers is meant the function $f(n)$ which is 1 or 0 according as n is or is not in the set.

Let u and v be fixed positive integers. Then, since

$$\sum_{j=1}^{u} \epsilon(nj/u)$$

is u or 0 according as n is or is not divisible by u, the characteristic function of the arithmetical progression

$$(71) \qquad u + v, \quad 2u + v, \cdots, \quad mu + v, \cdots$$

is represented by the trigonometric sum

$$(72) \qquad (1/u) \sum_{j=1}^{u} \epsilon(-vj/u)\epsilon(nj/u).$$

According to Legendre, the sums of three squares are the integers of the form $4^{k-1}(8m - 1)$, where $k = 1, 2, \cdots$ and $m = 1, 2, \cdots$. Let $f^k(n)$ denote the characteristic function of the subset which results when k is fixed and m varies. Then, if $f(n)$ is defined as at the beginning of (VIII), it is easy to prove that f is almost-periodic (B) and that, as a matter of fact, the k-th partial sum of the series $f^1 + f^2 + \cdots$ tends to f in the mean of the (B)-space, as $k \to \infty$ ([11], p. 27). Hence, corresponding to the transition from (17) to (14) in the proof of (I), the Fourier analysis (B) of f can be obtained by collecting the amplitudes which belong to a fixed frequency in the Fourier expansions of the k functions $f^1, \cdots, f^k$ and then letting $k \to \infty$. But this calculation can readily be carried out, as follows:

The set having the characteristic function f^k was defined to consist of the integers $4^{k-1}(8m - 1)$, where k is fixed and $m = 1, 2, \cdots$. Clearly, this set is identical with the case $u = 2^{2k+1}$, $v = -4^{k-1}$ of the arithmetical progression (71). Hence, the Fourier series (B) of $f(n)$ can be obtained by substituting $u = 2^{2k+1}$, $v = -4^{k-1}$ in (72) and then performing summation over all positive integers k. Consequently, the Fourier series (B) of $f(n)$ is

$$(73) \qquad f(n) \sim \sum_{k=1}^{\infty} \sum_{j=1}^{2^{2k+1}} 2^{-2k-1} \epsilon(j/8) \epsilon(nj/2^{2k+1}).$$

Since $\epsilon(n) = 1$, this means that there exist certain coefficients $b_r{}^q$ satisfying

$$(74) \qquad f(n) \sim \sum_{k=1}^{\infty} 2^{-2k-1} + \sum_{r=1}^{\infty} \sum_{1 \leqq q \leqq 2^r} b_r{}^q \, \epsilon(nq/2^r).$$

In order to prove (69) and (70), let r, q be two positive integers the second of which is odd and less than 2^r. Then $b_r{}^q$ in (74) is the sum of those coefficients $2^{-2k-1}\epsilon(j/8)$ of (73) for which the quotient $j/2^{2k+1}$ multiplying n has the fixed value $q/2^r$. Since $(q, 2^r) = 1$, these pairs of summation indices j, k can be parametrized by placing $j = qm$ and $2^{2k+1} = 2^r m$, where m runs through all positive integers. Consequently,

$$b_r{}^q = 2^{-r} \sum_{m=1}^{\infty} m^{-1} \epsilon(qm/8).$$

It follows therefore from (8) that (since the expansion $-\log(1-z) = \sum_{m=1}^{\infty} m^{-1} z^m$ remains valid on $|z| = 1$ if $z \neq 1$),

$$(75) \qquad b_r{}^q = -2^{-r} \log \{1 - \epsilon(q/8)\},$$

if the determination of the logarithm continuously is derived from that of $\log(1-z)$, where $|z| < 1$ and $\log 1 = 0$.

This completes the proof of (VIII), since (74) is identical with (69) in virtue of (75) and (70).

8. Theorems (III), (IV) and (VIII) have been deduced directly from (I), rather than from the complicated criterion (II). However, all that is responsible for the complications in (II) is the fact that the sufficient criterion supplied by (II) is necessary as well. Thus it will be convenient to isolate a *simple* particular case of (II), supplying a *sufficient* criterion which contains not only (III), (IV) and (VIII) but is also applicable, for instance, to the various sequences of Kloosterman's sums.

(IX) *Let* $e_1(n), e_2(n), \cdots$ *be a sequence of functions which are almost periodic* (B) *and have the following property: There exist constants*

$$(76) \qquad \alpha_r{}^s, \text{ where } (s, r) = 1,\ 1 \leqq s \leqq r \text{ and } r = 1, 2, \cdots,$$

such that the Fourier series (B) *of the functions* $e_1, e_2, \cdots$ *of* n *are expressible in terms of the sequence* $\beta_1, \beta_2, \cdots$ *of the trigonometric sums*

$$\beta_r(n) = \sum_{1 \leqq s \leqq r}^{(s,r)=1} \epsilon(-\alpha_r{}^s + ns/r) \tag{77}$$

in the form

$$e_r(n) \sim \sum_{d|r} \beta_d(n) \tag{78}$$

(which is equivalent to

$$\beta_r(n) \sim \sum_{d|r} \mu(r/d)\, e_d(n), \tag{79}$$

by Möbius' inversion). Then, if $g(n)$ *is any function satisfying*

$$\sum_{k=1}^{\infty} M(|\, e_k \,|)\, |\, g(k) \,| < \infty, \tag{80}$$

the function $f(n)$ *defined by the linear transformation*

$$f(n) = \sum_{m=1}^{n} e_m(n)\, g(n) \tag{81}$$

is almost-periodic (B) *and has the Fourier series* (14) *of the particular type*

$$f(n) \sim \sum_{r=1}^{\infty} b_r \beta_r(n), \qquad (B), \tag{82}$$

in which the ("shifted") Fourier constants b_r *are given by the absolutely convergent series*

$$b_r = \sum_{r|k} g(k). \tag{83}$$

In order to see this, let (81) be identified with (2). Then (3) is satisfied by $N_m = m$. On the other hand, due to the possibility allowed in (11), (7) and (9), it is clear from the device (47) applied in the proof of (IV), that the identification of (81) in (IX) with (2) in (I) is admissible (in (47), the assignments $n < m$ and $m \leqq n$ now become $n \leqq m$ and $m < n$ respectively). Hence, if $\lambda_m(k)$ in (30) is chosen to be 1 for every m and every k, it is sufficient to go over the assumptions and assertions of (II) step by step, in order to obtain (IX).

The formula corresponding to (41) in (III) is somewhat obscured in the generalization (IX) of (III), although (83) corresponds to (40). First, it is clear from (77) that only the first of the trigonometric sums $\beta_1(n)$, $\beta_2(n), \cdots$ contains a term independent of n, and that $\beta_1(n)$ is identical with this term, that is, with the constant $\epsilon(-\alpha_1{}^1)$. It follows therefore from (82) and (1) that $M(f)$, instead of being represented by the case $r = 1$ of (83), has the value

$$M(f) = \epsilon(-\alpha_1{}^1)\, b_1 = \epsilon(-\alpha_1{}^1) \sum_{k=1}^{\infty} g(k). \tag{84}$$

This also explains an apparent paradox in the wording of (IX), namely, the absolute convergence of the series (83) *whenever* (80) is satisfied. In fact, if all but one of the values $g(1), g(2), \cdots$ are chosen to be 0, it follows from (81) and (84) that

$$M(e_1) = M(e_2) = \cdot \cdot \cdot = \epsilon(-\alpha_1{}^1) \tag{85}$$

(this can be seen from (78) also). But $M(|h|)$ cannot be less than $|M(h)|$, and the exponential function ϵ occurring in (85) cannot become 0.

9. All of the arithmetical algorithms (2) considered thus far led to Fourier expansions (B) which, instead of having the *general* limit-periodic form (14), induce a "class condition," expressed by the phase rule (22) or by the corresponding reduction of (14) in (I) to (82), (77) in (IX). One might expect that this will be the case for every algorithm (2) of classical arithmetical interest. But it turns out that there is at least one classical algorithm, of substantially arithmetical significance, in which (14) cannot be contracted into (82), (77). It results if the "Eratosthenian" and "Eulerian" substitutions, (32) and (42), are replaced by the algorithm

$$f(n) = \sum_{m=1}^{n} \left\{ \frac{n}{m} - \left[\frac{n}{m}\right] \right\} f^0(m), \tag{86}$$

which underlies Dirichlet's divisor problem ([3]; cf. [4]).

In (86), the bracket [ ] is the symbol of the greatest integer not exceeding the quotient n/m. Hence, if (86) is thought of as written in the form (2), where

$$g(n) = f^0(n), \tag{87}$$

then (3) is satisfied by $N_m = m - 1$ for $m = 1, 2, \cdot \cdot \cdot$. It is also seen from (86) and (2) that, while

$$f(1) = 0, \qquad f(2) = 0, \tag{88}$$

$f(n)$ will contain $f^0(n-1)$ for every $n > 2$. In other words, if $n > 2$, then $f(n)$ is of the form $f(n) = F_{n-2} + e_{n-1}(n) f^0(n-1)$, where F_{n-2} is a linear form in the $n-2$ values $f^0(m)$ preceding $f^0(n-1)$, and the absolute constant $e_{n-1}(n)$ does not vanish. Consequently, there exists for every function $f(n)$ satisfying the initial restriction (88) a function $f^0(n)$ by means of which $f(n)$ is representable in the form (86) for every n, and f^0 is uniquely determined by f and by an arbitrary assignment of an initial value

$$f^0(1). \tag{89}$$

For every positive integer m, let $e^m = e^m(n)$ denote the non-negative function assigned by

$$(90)\qquad e^m(1) = 1/m, \quad e^m(2) = 2/m, \cdots, \quad e^m(m-1) = (m-1)/m; \quad e^m(m) = 0$$

and by the requirement that $e^m(n) = e^m(n+m)$ for every n (it is understood that, if $m = 1$, the assignment (90) gives

$$(91)\qquad e^1(n) = 0$$

for every n). In terms of these functions $e^m(n)$, define the functions $e_m(n)$ by (47). Then it is clear that (86) and (87) can be written in the form (2). Since the function e_m is non-negative and becomes identical with the periodic function e^m after the first period of e_m, it is also clear that $M(|e_m|) = M(e_m)$ exists and is identical with the arithmetical mean of the m values (90). Hence

$$M(|e_m|) = \sum_{j=1}^{m-1} j/m^2 \to \tfrac{1}{2}$$

as $m \to \infty$. Consequently, (5) is satisfied if and only if the series formed by the values (87) is absolutely convergent. This leads to the following theorem (in which the Fourier expansion (B) does *not* have the phase properties specified above):

(X) *For an arbitrary function* $f = f(n)$ *subject to* (88) *and for any choice of an initial constant* (89), *there exists a unique function* $f^0 = f^0(n)$ *by means of which* f *is representable in the form* (86). *Suppose that the function* f^0 *belonging to* f *satisfies the condition*

$$(92)\qquad \sum_{k=1}^{\infty} |f^0(k)| < \infty.$$

Then f *is almost-periodic* (B) *and has the Fourier series* (B)

$$(93)\qquad f(n) \sim \sum_{r=1}^{\infty} \sum_{\substack{(s,r)=1 \\ 1 \leq s \leq r}} a_r^s\, \epsilon(ns/r),$$

in which the Fourier constants are given by the absolutely convergent series

$$(94)\qquad a_r^s = \sum_{r|k} \lambda_r^s(k) f^0(k)/k^2,$$

where the λ *denote absolute constants*

$$(95)\qquad \lambda_r^s(k) = \sum_{n=1}^{k-1} n\epsilon(-ns/r).$$

The sum (95) denotes 0 when it is vacuous, that is, when $k = 1$; cf. (91).

First, the assumption (92) of (X) was seen to be identical with the assumption (5) of (I). Next, substitution of (90) into (13) gives

$$m^2\alpha_m(j) = \sum_{n=1}^{m-1} n\epsilon(-jn/m).$$

Hence, from (15) and (87),

$$a_r{}^s = \sum_{r|k} k^{-2}f^0(k) \sum_{n=1}^{k-1} n\epsilon(-(ksn/k)/r).$$

But this can be written in the form (94), if $\lambda_r{}^s(k)$ is defined by (95).

Accordingly, (X) follows by applying (I) to the case of Dirichlet's algorithm.

10. If $1 \leqq m \leqq n$ and $n = 1, 2, \cdots$, let $r_n(m)$ denote the residue which remains when n is divided by m. In view of the description of the matrix of (86) in **9**, it is natural to raise the question as to the asymptotic behavior of the *average fluctuation* in the set

$$r_n(1),\ r_n(2), \cdots,\ r_n(m), \cdots,\ r_n(n) \tag{96}$$

of the subsequent division residues of n, as $n \rightarrow \infty$. For a fixed n, this average fluctuation is meant to be the arithmetical mean of the *absolute* deviations of the *consecutive* elements in the ordered set (96) of n non-negative integers, that is, the arithmetical mean

$$\phi_n = (1/n) \sum_{m=1}^{n} | r_n(m+1) - r_n(m) |. \tag{97}$$

The value of $r_n(m+1)$ for $m = n$ in (97) is, in the main, immaterial as $n \rightarrow \infty$; it can naturally be defined by

$$r_n(n+1) = n, \tag{98}$$

since the least non-negative residue $(\operatorname{mod} n+1)$ of n is n.

An answer to the question is contained in the following fact:

The average fluctuation of the consecutive division residues (96) *of n is of the order of* $\log n$; *in the sense that, if ϕ_n is defined by* (97),

$$0 < \liminf_{n\rightarrow\infty} \phi_n/\log n \leqq \limsup_{n\rightarrow\infty} \phi_n/\log n < \infty. \tag{99}$$

It will remain undecided whether or not this answer is of a final nature, since the approach to be followed will neither prove nor disprove the possibility that the (positive and finite) lower and upper limits coincide in (99).

As recently proved in another connection ([11], pp. 11-12),

$$\sum_{m=1}^{n} \left| \left[\frac{n}{m}\right]\frac{m}{n} - \left[\frac{n}{m+1}\right]\frac{m+1}{n} \right| > \tfrac{1}{2}\log n - \text{const.}$$

On the other hand, if $e_m(n)$ denotes, as in **9**, the coefficient of $f^0(m)$ in the n-th of the equations (86), then, since $e_m(n) = n/m - [n/m]$, it is clear that

$$me_m(n) - (m+1)e_{m+1}(n) = m[n/m] - (m+1)[n/(m+1)].$$

But it is also clear from $e_m(n) = n/m - [n/m]$ that

$$r_n(m) = me_m(n),$$

since $[n/m]$ denoted the integral part of the quotient n/m, while $r_m(n)$ was defined as the least non-negative residue of n (mod m). The last three formula lines imply the first of the inequalities (99), since ϕ_n is defined by (97).

In the direction of the remaining assertion of (99), it is clear from the last two formula lines that

$$|r_n(m) - r_n(m+1)| \leqq |[n/m] - [n/(m+1)]|\, m + [n/m].$$

Since $[n/(m+1)]$ is non-negative and does not exceed $[n/m]$, the sign of absolute value is superfluous on the right of this inequality. Consequently, summation with respect to m gives

$$(100) \qquad n\phi_n \leqq \sum_{m=1}^{n} [n/m]m - \sum_{m=1}^{n} [n/(m+1)]m + \sum_{m=1}^{n} [n/m],$$

ϕ_n being defined by (97). If m is replaced by $m-1$, the second of the three terms on the right of (100) appears in the form

$$-\sum_{m=2}^{n+1} [n/m](m-1),$$

which, since $[n/m](m-1)$ vanishes for $m = 1$ and for $m = n+1$, is identical with

$$-\sum_{m=1}^{n} n/m(m-1), \text{ i.e., } -\sum_{m=1}^{n} [n/m]m + \sum_{m=1}^{n} [n/m].$$

Accordingly, (100) can be contracted into

$$n\phi_n \leqq 2\sum_{m=1}^{n} [n/m]; \text{ so that } \phi_n \leqq 2\sum_{m=1}^{n} 1/m,$$

since $[n/m]/n \leqq 1/m$. It follows therefore from $\sum_{m=1}^{n} 1/m \sim \log n$ that the proof of the last of the inequalities (99) is complete.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] A. S. Besicovich, *Almost Periodic Functions*, Cambridge, 1932.

[2] R. D. Carmichael, "Expansions of arithmetical functions in infinite series," *Proceedings of the London Mathematical Society*, Ser. 2, vol. 34 (1932), pp. 1-26.

[3] G. P. L. Dirichlet, "Ueber die Bestimmung der mittleren Werthe in der Zahlentheorie," *Werke*, vol. 2, pp. 49-66.

[4] ———, "Ueber ein die Division betreffendes Problem," *Werke*, vol. 2, pp. 97-104.

[5] G. H. Hardy, *Ramanujan*, Cambridge, 1940.

[6] ———, "Eratosthenian averages," *Nature*, vol. 152 (1943), p. 708.

[7] O. Hölder, "Zur Theorie der Kreisteilungsgleichung $K_m(x) = 0$," *Prace Matematyczno-Fizyczne*, vol. 43 (1936), pp. 13-23.

[8] G. Libri, "Mémoire sur la théorie des nombres," *Journal für die reine und angewandte Mathematik*, vol. 9 (1832), pp. 54-80.

[9] S. Ramanujan, "On certain trigonometrical sums and their applications in the theory of numbers," *Collected Papers*, pp. 178-199.

[10] H. Tegnér, "Von dem Sylvesterschen Denumeranten," *Arkiv för Mat., Astr. och Fys.*, vol. 23 A (1932), no. 7.

[11] A. Wintner, *Eratosthenian Averages*, Baltimore, 1943.

[12] ———, *The Theory of Measure in Arithmetical Semi-Groups*, Baltimore, 1944.

A PARTITION FUNCTION WITH THE PRIME MODULUS $P > 3$.*

By JOHN LIVINGOOD.**

1. Recently Lehner [2][1] developed a convergent series for $p_1(n)$ and $p_2(n)$, the number of partitions of a positive integer n into summands of the form $5l \pm 1$ and $5l \pm 2$ respectively. The purpose of this paper is to develop a similar series for $p_1(n), p_2(n), \cdots, p_{(p-1)/2}(n)$, the number of partitions of a positive integer n into summands of the form $pl \pm 1$, $pl \pm 2, \cdots$, $pl \pm (p-1)/2$, p being a prime greater than 3. Here, as in the previous paper, we follow the method of Rademacher [5].

We consider the generating functions

$$F_a(x) = \prod_{l=1}^{\infty} (1 - x^l)^{-1} = \sum_{n=0}^{\infty} p_a(n) x^n \tag{1.1}$$

$$l \equiv \pm a \pmod{p}, \qquad (a = 1, 2, \cdots, (p-1)/2)$$

convergent within the unit circle. To determine the asymptotic behavior of $F_a(x)$ near a rational point on a circle concentric to the unit circle but interior to it, we subject x to the transformation $x \rightarrow x'$ where

$$x = \exp(2\pi i h/k - 2\pi z/k), \quad x' = \exp(2\pi i h'/k - 2\pi/kz). \tag{1.2}$$

Here $\mathcal{R}(z) > 0$, h and k are coprime integers satisfying $0 \leqq h < k$, and h' is any fixed solution of

$$hh' \equiv -1 \pmod{k}. \tag{1.21}$$

We then derive a functional equation connecting $F_a(x)$ and $F_b(x')$ for all k divisible by p ($b = 1, 2, \cdots, (p-1)/2$ depending upon h). If $p \nmid k$, we write

$$x'' = \exp(2\pi i H'/k - 2\pi/Kz), \qquad HH' \equiv -1 \pmod{k} \tag{1.3}$$

and find a similar relation between $F_a(x)$ and certain new functions $H_a(x'')$. Certain sums of roots of unity appear. The application of the Rademacher method requires that these sums be reduced to incomplete Kloosterman sums, so that a better estimate than $O(k)$ will result.

---

* Received April 24, 1944.

** The author wishes to express his appreciation to Professor Hans Rademacher for suggesting this problem and for giving much helpful advice in its solution.

[1] Square bracket numbers refer to the bibliography.

We now introduce the Cauchy integral and apply the transformation equations (3.93) and (3.94). If $p \mid k$ we obtain

$$
\begin{aligned}
&p_a^{(1)}(n) = \sum_{b=1}^{(p-1)/2} {\sum_{h,k}}' \omega_a(h,k) \exp(-2\pi i h n/k) \\
(1.4) \qquad &\times \int_{-\theta'}^{\theta''} \sum_{\nu=0}^{\infty} p_b(\nu) \exp(2\pi i h'\nu/k) \\
&\times \exp\{-(\pi/k^2 w)(2\nu - B/6p) + \pi w(2n - A/6p)\} d\phi, \\
&0 \leq h < k \leq N, \quad \theta' = \theta'_{h,k}, \quad \theta'' = \theta''_{h,k}, \quad B = p^2 - 6pb + 6b^2.
\end{aligned}
$$

Next the integrand is split into two parts, depending upon the sign of the coefficient of w^{-1}. We note that this coefficient is positive for $\nu = 0, 1, \cdots, [B/12p]$. For certain p, $\nu = 0$ is the only admissible value. This is the result that Lehner obtained for $p = 5$. It is also true for all $p \leq 17$. However, for other p, there may be more admissible values for ν. For example, if $p = 19$, $\nu = 0$ or 1. This interesting information results in the extension of the previous work. The contributions furnished by the case $p \mid k$ when the coefficient of w^{-1} is negative, and by the case $p \nmid k$, follow as in the special case $p = 5$.

Finally we obtain asymptotic ratios for $p_1(n), p_2(n), \cdots, p_{(p-1)/2}(n)$.

I. The Transformation Equations.

2. The first object is to derive a transformation equation for $F_a(x)$ in (1.1). Consider the case $p \mid k$. $F_a(x)$ is regular, without zeros, in the unit circle; therefore $\log F_a(x)$ is single-valued in the same region if we choose a specific branch of the logarithm, for example, the one given by $\log F(0) = 0$. Then

$$
(2.1) \qquad G_a(x) = \log F_a(x) = -\sum_{l=1}^{\infty} \log(1 - x^l) = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} (x^{lm}/m), \qquad l \equiv \pm a \pmod{p}.
$$

Placing

$$
\begin{aligned}
&x = \exp(2\pi i h/k) \times X, \quad X = \exp(-y), \quad \mathcal{R}(y) > 0 \\
(2.2) \qquad &m = qk + \mu, \quad 1 \leq \mu \leq k; \qquad q, r = 0, 1, 2, \cdots \\
&l = rk + \lambda, \quad 1 \leq \lambda < k, \quad \lambda \equiv \pm a \pmod{p}
\end{aligned}
$$

in (2.1) we have

$$
(2.21) \qquad G_a(x) = \sum_{\lambda} \sum_{\mu} \exp(2\pi i h \mu \lambda/k) \sum_{r,q} \frac{1}{qk + \mu} \exp\{-(rk + \lambda)(qk + \mu) y\}
$$

where λ, μ, r, and q assume the values described above.[2]

---

[2] Throughout this paragraph, λ, μ, q and r assume these values.

Application of the Mellin formula to (2.21) gives

$$G_a(x) = \sum_{\lambda}\sum_{\mu} \exp(2\pi ih\mu\lambda/k) \sum_{r,q} (1/2\pi i) \int_{(\alpha)} \frac{\Gamma(s)}{qk+\mu} \; \frac{ds}{(rk+\lambda)^s(qk+\mu)^s y^s}$$

$$= \sum_{\lambda}\sum_{\mu} \exp(2\pi ih\mu\lambda/k)\,(1/2\pi i) \int_{(\alpha)} \frac{\Gamma(s)}{y^s k^{2s+1}} \; \sum_{r=0}^{\infty} (r+\lambda/k)^{-s} \sum_{q=0}^{\infty} (q+\mu/k)^{-s-1} ds$$

$$(2.3) \qquad G_a(x) = \frac{1}{2\pi ik} \; \sum_{\lambda}\sum_{\mu} \exp(2\pi ih\mu\lambda/k) \int_{(\alpha)} \frac{\Gamma(s)}{y^s k^{2s}} \zeta(s, \lambda/k)\zeta(1+s, \mu/k)\,ds$$

where

$$\zeta(s,a) = \sum_{n=0}^{\infty} \frac{1}{(n+a)^s}, \qquad 0 < a \leq 1$$

is the Hurwitz Zeta function, and where the (α) under the integral sign means that the path of integration is from $\alpha - i\infty$ to $\alpha + i\infty$.

Recall the Zeta functional equation

$$(2.4) \qquad \begin{aligned} \zeta(s, \lambda/k) = \Gamma(1-s) \frac{2}{(2\pi k)^{1-s}} \{&\sin(\pi s/2) \sum_{\nu=1}^{k} \cos 2\pi\lambda\nu/k \cdot \zeta(1-s, \nu/k) \\ &+ \cos(\pi s/2) \sum_{\nu=1}^{k} \sin 2\pi\lambda\nu/k \cdot \zeta(1-s, \nu/k)\}. \end{aligned}$$

Applying (2.4), replacing[3] $\exp(2\pi ih\mu\lambda/k)$ by $\cos(2\pi h\mu\lambda/k)$ and $\exp(2\pi ih\mu\lambda/k)$ by $i\sin(2\pi h\mu\lambda/k)$ in the first and second terms of the right member of (2.3), and setting $z = (yk/2\pi)$, we have, after simplification

$$(2.41) \qquad \begin{aligned} G_a(x) = {} & \frac{1}{4\pi ik^2} \sum_{\lambda}\sum_{\mu,\nu} \cos 2\pi h\mu\lambda/k \cos 2\pi\lambda\nu/k \\ & \times \int_{(\alpha)} \frac{\zeta(1+s, \mu/k)\zeta(1-s, \nu/k)}{z^s \cos \pi s/2} ds \\ & + \frac{1}{4\pi k^2} \sum_{\lambda}\sum_{\mu,\nu} \sin 2\pi h\mu\lambda/k \sin 2\pi\lambda\nu/k \\ & \times \int_{(\alpha)} \frac{\zeta(1+s, \mu/k)\zeta(1-s, \nu/k)}{z^s \sin \pi s/2} ds \end{aligned}$$

where $z^s = \exp(s \log z)$ and where we can take $|\mathfrak{J} \log z| < \pi/2$ since $\mathfrak{R}(z) > 0$.

We now define ρ by setting

$$(2.5) \qquad \lambda \equiv h'\rho \pmod{k}, \qquad 0 < \rho < k$$

where h' is given by (1.21). This, together with (1.21), shows that

$$(2.51) \qquad \rho \equiv -h\lambda \pmod{k}.$$

---

[3] This is permissible since we sum over λ and since $\lambda \equiv \pm a \pmod{p}$.

In addition, since $p \mid k$, (1.21) implies $hh' \equiv -1 \pmod{p}$ and hence, by Fermat's Theorem,

$$h' \equiv -h^{p-2} \pmod{p}.$$

Since $p \mid k$ and $\lambda \equiv \pm a \pmod{p}$, (2.51) implies

$$\rho \equiv \pm ha \equiv \pm b \pmod{p}. \tag{2.52}$$

Introducing the new summation letter ρ and replacing s by $-s$, we have

$$\begin{aligned} G_a(x) = {} & (1/4\pi i k^2) \sum_{\rho} \sum_{\mu,\nu} \cos 2\pi\mu\rho/k \cos 2\pi h'\rho\nu/k \\ & \times \int_{(-\alpha)} \frac{\zeta(1-s,\mu/k)\zeta(1+s,\nu/k)}{z^{-s}\cos \pi s/2}\, ds \\ & + (1/4\pi k^2) \sum_{\rho} \sum_{\mu,\nu} \sin 2\pi\mu\rho/k \sin 2\pi h'\rho\nu/k \\ & \times \int_{(-\alpha)} \frac{\zeta(1-s,\mu/k)\zeta(1+s,\nu/k)}{z^{-s}\sin \pi s/2}\, ds \end{aligned} \tag{2.6}$$

where the restrictions on ρ are given by (2.5) and (2.52).

The appearance of $\cos \pi s/2$ and $\sin \pi s/2$ in the denominators enables us to shift the path of integration from $(-\alpha)$ to (α), taking into consideration the residues at the various poles. Doing this and comparing the result with (2.41), we find two expressions of the same form, except that we now have ρ instead of λ, h' instead of h, μ and ν interchanged, and z^{-1} instead of z. Hence, using the definition (1.2) of x', we obtain

$$G_a(x) = G_b(x') - 2\pi i(R_1 + R_2) \tag{2.7}$$

where R_1 and R_2 are the sums of the residues of the first and second terms respectively in the right member of (2.6) with the sign of integration moved to the extreme left in each case.

These residues turn out to be the following:[4]

$$\begin{gathered} R_1 = Az/12ikp - B/12ikpz; \\ R_2 = -\tfrac{1}{2} \sum_{\lambda=1}^{k-1} ((\lambda/k))((h\lambda/k)), \qquad \lambda \equiv \pm a \pmod{p} \end{gathered} \tag{2.8}$$

where

$$\begin{gathered} A = p^2 - 6pa + 6a^2, \qquad B = p^2 - 6pb + 6b^2, \\ ((x)) = x - [x] - \tfrac{1}{2} + \tfrac{1}{2}\delta(x). \end{gathered} \tag{2.81}$$

Thus, by use of (2.7) and (2.8) and by taking exponentials, we obtain the desired transformation equation for the case $p \mid k$:

---

[4] Calculation of the residues follows as in Lehner's paper.

$$F_a(x) = \omega_a(h,k)\exp\{(\pi/6kp)(B/z - Az)\}F_b(x'). \tag{2.9}$$

Here x and x' are defined by (1.2), b by (2.52), A and B by (2.81) and

$$\omega_a(h,k) = \exp\{\pi i\sigma_a(h,k)\} \tag{2.91}$$

with

$$\sigma_a(h,k) = \sum_{\lambda=1}^{k-1}((\lambda/k))((h\lambda/k)), \qquad \lambda \equiv \pm a \pmod{p}. \tag{2.92}$$

3. Now we consider the case $p \nmid k$. Here $F_a(x)$ does not transform into $F_b(x')$; however, $F_a(x)$ will be transformed into a certain new function by a method similar to the preceding one. Choose x'', H, H', K as in (1.3) and place

$$\begin{aligned} &x = \exp(2\pi ih/k) \times X, \qquad X = \exp(-y), \qquad \mathfrak{R}(y) > 0 \\ &m = qk + \mu, \qquad 1 \leq \mu \leq k; \qquad r, q = 0, 1, 2, \cdots \\ &l = rK + \lambda, \qquad 1 \leq \lambda < K, \qquad \lambda \equiv \pm a \pmod{p}. \end{aligned} \tag{3.1}$$

As in **2**, we find[5]

$$\begin{aligned} G_a(x) = {} & (1/4\pi ikK)\sum_\lambda\sum_\mu\sum_\nu \cos 2\pi h\mu\lambda/k \cos 2\pi\lambda\nu/K \\ & \times \int_{(a)} \frac{\zeta(1+s,\mu/k)\zeta(1-s,\nu/K)}{z^s\cos\pi s/2}\,ds \\ & + (1/4\pi kK)\sum_\lambda\sum_\mu\sum_\nu \sin 2\pi h\mu\lambda/k \sin 2\pi\lambda\nu/K \\ & \times \int_{(a)} \frac{\zeta(1+s,\mu/k)\zeta(1-s,\nu/K)}{z^s\sin\pi s/2}\,ds; \\ & \qquad\qquad (\nu = 1, 2, \cdots, K). \end{aligned} \tag{3.2}$$

Define λ^* by setting

$$\lambda \equiv pH'\lambda^* \pmod{k}, \qquad 0 < \lambda^* \leq k \tag{3.3}$$

where H' is any fixed solution of $HH' \equiv -1 \pmod{k}$ and $H = ph$. We observe that (3.3) implies $\lambda^* \equiv -h\lambda \pmod{k}$. Introducing λ^*, replacing s by $-s$, and shifting the path of integration as in **2**, we obtain

$$\begin{aligned} G_a(x) = {} & (1/4\pi ikK)\sum_\lambda\sum_\mu\sum_\nu \cos 2\pi\mu\lambda^*/k \cos 2\pi\lambda\nu/K \\ & \times \int_{(a)} \frac{\zeta(1-s,\mu/k)\zeta(1+s,\nu/K)}{z^{-s}\cos\pi s/2}\,ds \\ & + (1/4\pi kK)\sum_\lambda\sum_\mu\sum_\nu \sin 2\pi\mu\lambda^*/k \sin 2\pi\lambda\nu/K \\ & \times \int_{(a)} \frac{\zeta(1-s,\mu/k)\zeta(1+s,\nu/K)}{z^{-s}\sin\pi s/2}\,ds \\ & - 2\pi i(R_1 + R_2) = L_a(x) - 2\pi i(R_1 + R_2). \end{aligned} \tag{3.4}$$

---

[5] Throughout this paragraph, λ and μ assume the values given in (3.1) and ν assumes the values $1, 2, \cdots, K$.

In order to obtain the new function into which $F_a(x)$ is transformed, we now retrace our steps from (3. 4). Replace $\cos(2\pi\lambda\nu/K)$ by $\exp(2\pi i\lambda\nu/K)$ and $\sin(2\pi\lambda\nu/K)$ by $i^{-1}\exp(2\pi i\lambda\nu/K)$ and apply the transformation equation for $\zeta(s, \lambda^*/k)$. After simplification, we find

$$L_a(x) = (1/2\pi iK)\sum_{\lambda}\sum_{\nu}\exp(2\pi i\lambda\nu/K)$$
$$\times\int_{(a)}(z^s\Gamma(s)/(2\pi k)^s)\zeta(s, \lambda^*/k)\zeta(1+s, \nu/K)\,ds.$$

Application of the Mellin formula gives

$$L_a(x) = \sum_{\lambda}\sum_{\nu}\exp(2\pi i\lambda\nu/K)\sum_{r,q=0}^{\infty}(rK+\nu)^{-1}\exp\{-(2\pi/Kz)(qk+\lambda^*)(rK+\nu)\}$$
(3. 5)
$$= -\sum_{\lambda}\sum_{q}\log\{1-\exp[-(2\pi/Kz)(qk+\lambda^*)+2\pi i\lambda/K]\}.$$

Examination of (3. 1) reveals that since $p\nmid k$, λ runs over a complete residue system modulo k twice in some order. We therefore define α by

(3. 6) $$k\alpha \equiv a \pmod{p}, \qquad 0 < \alpha < p.$$

This, along with (3. 3), yields

(3. 61) $$\lambda \equiv pH'\lambda^* \pm k\alpha \pmod{K}.$$

Inserting (3. 61) into (3. 5), we obtain

$$L_a(x) = -\sum_{\lambda}\sum_{q}\log\{1-\exp[-(2\pi/Kz)(qk+\lambda^*)$$
$$+2\pi iH'(qk+\lambda^*)/k \pm 2\pi i\alpha/p]\}.$$

Place

(3. 71) $$qk+\lambda^* = l, \exp(2\pi i\alpha/p) = \eta(\alpha), \exp(-2\pi i\alpha/p) = \overline{\eta(\alpha)}.$$

Using these notations and (1. 3), we find

$$L_a(x) = -\sum_{l=1}^{\infty}\log(1-\eta(\alpha)x''^{l}) - \sum_{l=1}^{\infty}\log(1-\overline{\eta(\alpha)}x''^{l})$$

(3. 72) $$L_a(x) = \log\prod_{l=1}^{\infty}(1-\eta(\alpha)x''^{l})^{-1}\prod_{l=1}^{\infty}(1-\overline{\eta(\alpha)}x''^{l})^{-1} = \log H_a(x'')$$

where

(3. 73) $$H_a(x) = \prod_{l=1}^{\infty}(1-\eta(\alpha)x^{l})^{-1}\prod_{l=1}^{\infty}(1-\overline{\eta(\alpha)}x^{l})^{-1} = 1+\sum_{\nu=1}^{\infty}C_a(\nu)x^{\nu}.$$

We now calculate the residues as in Lehner's paper, and find

$$R_1 = Az/12ikp - 1/12ikpz + \tfrac{1}{2}\log(2\sin\pi\alpha/p);$$
(3. 8)
$$R_2 = -\tfrac{1}{2}\sum_{\lambda=1}^{K-1}((h\lambda/k))((\lambda/K)), \quad \lambda \equiv \pm a \pmod{p}$$

where A and $((x))$ are given by (2. 81). Therefore, by use of (3. 4), (3. 72) and (3. 8) and by taking exponentials, we obtain the desired transformation equation for the case $p \nmid k$:

$$(3.9) \quad F_a(x) = \tfrac{1}{2}\chi_a(h, k) \csc \pi\alpha/p \exp \{(\pi/6kp)(1/z - Az)\} H_a(x'')$$

where A is given by (2. 81), $H_a(x)$ by (3. 73), x by (1. 2), x'' by (1. 3) and

$$(3.91) \qquad \chi_a(h, k) = \exp \{\pi i t_a(h, k)\}$$

with

$$(3.92) \qquad t_a(h, k) = \sum_{\lambda=1}^{K-1} ((h\lambda/k))((\lambda/K)), \qquad \lambda \equiv \pm a \pmod{p}.$$

From (2. 9) and (3. 9) we have the following result:

THEOREM 1. *The function $F_a(x)$, defined in (1. 1), satisfies the transformation equation*

$$\begin{aligned}(3.93) \qquad & F_a(\exp [2\pi i h/k - 2\pi z/k]) \\ & = \omega_a(h, k) \exp [(\pi/6kp)(B/z - Az)] F_b(\exp [2\pi i h'/k - 2\pi/kz])\end{aligned}$$

for the case $p | k$, and the equation

$$\begin{aligned}(3.94) \qquad & F_a(\exp [2\pi i h/k - 2\pi z/k]) \\ & = \tfrac{1}{2}\chi_a(h, k) \csc \pi\alpha/p \exp [(\pi/6kp)(1/z - Az)] H_a(exp\ [2\pi i H'/k - 2\pi/K\end{aligned}$$

for the case $p \nmid k$.

II. The Kloosterman Sums.

4. When the transformation equations (3. 93) and (3. 94) are inserted into the Cauchy integral, certain sums of the roots of unity $\omega_a(h, k)$ and $\chi_a(h, k)$ appear. The estimate $O(k)$ for these sums is not sufficient for the application of Rademacher's method. Thus, we now reduce these sums to incomplete Kloosterman sums, subject to the estimate $O(n^{1/3}k^{2/3+\epsilon})$, according to Salié.

Consider the case $p | k$. From (2. 92)

$$\sigma_a(h, k) = \sum_{\lambda=1}^{k-1} ((\lambda/k))((h\lambda/k)), \qquad \lambda \equiv \pm a \pmod{p}.$$

However, by the same procedure as that used by Lehner,[6] we are able to replace this by

$$(4.1) \quad \sigma_a(h, k) = 2 \sum_{\lambda=1}^{k-1} (\lambda/k)((h\lambda/k)) - (b/p - \tfrac{1}{2}), \qquad \lambda \equiv a \pmod{p}$$

where b is given by (2. 52). Employing the definition of $((x))$ we obtain

---

[6] See [2], pages 644-645.

$$\sigma_a(h,k) = (2h/k^2)\sum_\lambda \lambda^2 - (2/k)\sum_\lambda \lambda[h\lambda/k] - (1/k)\sum_\lambda \lambda - (b/p - \tfrac{1}{2}), \qquad \lambda \equiv a \pmod{p}$$

which upon simplification becomes

$$\begin{aligned} 6pk\sigma_a(h,k) &= 2h\{2k^2 + 3k(2a - p) + A\} \\ &\quad - 3k(k - 2p + 2a + 2b) - 12p\sum_\lambda \lambda[h\lambda/k], \qquad \lambda \equiv a \pmod{p} \end{aligned} \tag{4.2}$$

where A is given by (2.81). Examination of (4.2) reveals that $6pk\sigma_a(h,k)$ is always integral. Also,

$$\begin{aligned} 6pk\sigma_a(h,k) &\equiv 0 \pmod{3} \text{ if } 3\nmid k, \\ 6pk\sigma_a(h,k) &\equiv 2a + 2b + 2p - 3 \pmod{4} \text{ if } k \text{ is odd.} \end{aligned} \tag{4.21}$$

Next we shall evaluate a certain sum in two different ways,[7] and obtain a result containing $\sigma_a(h,k)$, from which we can find the residues of $6pk\sigma_a(h,k)$ for moduli which are multiples of k. Now

$$\sum_\lambda ((h\lambda/k))^2 = \sum_\rho ((\rho/k))^2 = \sum_\rho (\rho/k - \tfrac{1}{2})^2 = (B - k^2)/6pk + \tfrac{1}{4}\sum_\lambda 1; \qquad \lambda \equiv a \pmod{p},\quad \rho \equiv -b \pmod{p}. \tag{4.31}$$

Also

$$\begin{aligned} \sum_\lambda ((h\lambda/k))^2 &= \sum_\lambda (h\lambda/k - [h\lambda/k] - \tfrac{1}{2})^2 \\ &= h\sigma_a(h,k) + h(b/p - \tfrac{1}{2}) - (h^2/6pk)\{2k^2 + 3k(2a - p) + A\} \\ &\quad + \sum_\lambda [h\lambda/k]([h\lambda/k] + 1) + \tfrac{1}{4}\sum_\lambda 1; \qquad \lambda \equiv a \pmod{p} \end{aligned} \tag{4.32}$$

by the use of (4.1). Comparing (4.31) and (4.32) we find

$$\begin{aligned} 6phk\sigma_a(h,k) &= h^2\{2k^2 + 3k(2a - p) + A\} - (k^2 - B) \\ &\quad - 3hk(2b - p) - 12pkT \end{aligned} \tag{4.4}$$

where

$$T = \tfrac{1}{2}\sum_\lambda [h\lambda/k]([h\lambda/k] + 1), \qquad \lambda \equiv a \pmod{p}$$

is always integral. Let $12p = fG$, where f is the greatest divisor of $12p$ which is prime to k. We obtain the following results:

$$\begin{array}{llll} (k,12p) = p, & f = 12,\ G = p & (k,12p) = 3p, & f = 4,\ G = 3p \\ \left.\begin{array}{l}(k,12p) = 2p \\ (k,12p) = 4p\end{array}\right\}, & f = 3,\ G = 4p & \left.\begin{array}{l}(k,12p) = 6p \\ (k,12p) = 12p\end{array}\right\}, & f = 1,\ G = 12p. \end{array} \tag{4.5}$$

Denote by h' a solution of $hh' \equiv -1 \pmod{Gk}$. Since all primes in $G|k$, this congruence always has solutions because $(h,k) = 1$ implies $(h, Gk) = 1$.

Multiplying (4.4) by $-h'$ and remembering that $Gk|2k^2$, we obtain

---

[7] See [6], proof of Theorem 3.

$$6pk\sigma_a(h,k) \equiv hu - h'v - 3k(2b-p) \pmod{Gk} \tag{4.61}$$

where

$$u = 3k(2a-p) + A, \qquad v = B - k^2.$$

From (4.21) and (4.5) we have

$$6pk\sigma_a(h,k) \equiv 6a + 6b + 6p - 3 \pmod{f}. \tag{4.62}$$

Set

$$f\phi \equiv 1 \pmod{Gk}, \qquad Gk\Gamma \equiv 1 \pmod{f}. \tag{4.7}$$

Then, from (4.61) and (4.62) we find

$$\begin{aligned} 6pk\sigma_a(h,k) \equiv{}& f\phi\{hu - h'v - 3k(2b-p)\} \\ &+ Gk\Gamma(6a + 6b + 6p - 3) \pmod{12kp}. \end{aligned} \tag{4.8}$$

Thus, finally,

$$\begin{aligned} \omega_a(h,k) &= \exp\ \{\pi i\sigma_a(h,k)\} = \exp\{(2\pi i/12kp)6pk\sigma_a(h,k)\} \\ &= \exp\{2\pi i[(\phi/Gk)(hu - h'v) - (3\phi/G)(2b-p) \\ &\quad + (\Gamma/f)(6a + 6b + 6p - 3)]\}. \end{aligned} \tag{4.9}$$

5. We now consider the case $p \nmid k$. From (3.92) and an argument similar to that of Lehner's, we have

$$\begin{aligned} t_a(h,k) &= 2\sum_{\lambda=1}^{K-1}(\lambda/K)((h\lambda/k)) = 2\sum_{\lambda=1}^{K-1}(\lambda/K)(h\lambda/k - [h\lambda/k] \\ &\quad - \tfrac{1}{2} + \tfrac{1}{2}\delta(h\lambda/k)), \ \lambda \equiv a \pmod{p}. \end{aligned} \tag{5.1}$$

As in **4**, we obtain, upon simplification,

$$\begin{aligned} 6pkt_a(h,k) ={}& 2h\{2K^2 + 3K(2a-p) + A\} \\ &- 3k(K - p + 2a - 2\alpha) - 12\sum_{\lambda}\lambda[h\lambda/k], \\ &\lambda \equiv a \pmod{p} \end{aligned} \tag{5.2}$$

where α is given by (3.6). Examination of (5.2) shows

$$\begin{aligned} 6pkt_a(h,k) &\equiv 0 && \pmod{3} \text{ if } 3 \nmid k \\ 6pkt_a(h,k) &\equiv (p - pk + 2a + 2\alpha) && \pmod{4} \text{ if } k \text{ is odd} \\ 6pkt_a(h,k) &\equiv 0 && \pmod{p}. \end{aligned} \tag{5.3}$$

As in **4**, we evaluate $\sum_{\lambda}((h\lambda/k))^2$, $\lambda \equiv a \pmod{p}$, $\lambda = 1, 2, \cdots, K-1$

in two different ways, and upon comparing results find

$$\begin{aligned} 6pkt_a(h,k) ={}& h^2\{2K^2 + 3K(2a-p) + A\} \\ &- (k^2 - 1) + 6hk\alpha - 12kT' \end{aligned} \tag{5.4}$$

where

$$T' = \tfrac{1}{2}\sum_{\lambda}[h\lambda/k]([h\lambda/k] + 1), \qquad \lambda \equiv a \pmod{p}.$$

Let $12p = Fg$, where F is the greatest divisor of $12p$ which is prime to k. Then

$$(5.5)\quad \begin{array}{lll} (k,12p)=1, & F=12p,\ g=1 & \qquad (k,12p)=3, \quad F=4p,\ g=3 \\ \left.\begin{array}{l}(k,12p)=2\\(k,12p)=4\end{array}\right\}, & F=3p,\ g=4 & \qquad \left.\begin{array}{l}(k,12p)=6\\(k,12p)=12\end{array}\right\}, \quad F=p,\ g=12. \end{array}$$

Denote by h' a solution of $hh' \equiv -1 \pmod{gk}$. Multiplying (5.4) by $-h'$ and remembering that $g \mid 12$ and $g \mid 2k$, we obtain

$$(5.61)\qquad 6pkt_a(h,k) \equiv hu - h'v + 6k\alpha \pmod{gk}$$

where

$$u = 3K(2a-p) + A, \qquad v = 1 - k^2.$$

From (5.3) and (5.5)

$$(5.62)\qquad 6pkt_a(h,k) \equiv 9p^2(2a + 2\alpha + p - pk) \pmod{F}.$$

Setting

$$(5.7)\qquad F\Phi \equiv 1 \pmod{gk}, \qquad gk\gamma \equiv 1 \pmod{F}$$

and combining (5.61) and (5.62), we find

$$(5.8)\qquad \begin{aligned} 6pkt_a(h,k) \equiv{} & F\Phi(hu - h'v + 6k\alpha) \\ & + 9gk\gamma p^2(2a + 2\alpha + p - pk) \pmod{12pk}. \end{aligned}$$

Therefore

$$\begin{aligned} \chi_a(h,k) &= \exp\{\pi i t_a(h,k)\} = \exp\{(2\pi i/12pk)6pkt_a(h,k)\} \\ &= \exp\{2\pi i[(\Phi/gk)(hu - h'v) + 6\Phi\alpha/g + (9p^2\gamma/F)(2a + 2\alpha + p - pk)]\}. \end{aligned}$$

By choosing $pH' \equiv h' \pmod{gk}$, and using $G = pg$, $H = ph$ we obtain finally

$$(5.9)\qquad \begin{aligned} \chi_a(h,k) = \exp\{2\pi i[&(\Phi/Gk)(Hu - p^2H'v) \\ & + 6\Phi\alpha/g + (9p^2\gamma/F)(2a + 2\alpha + p - pk)]\}. \end{aligned}$$

6. Consider the sum

$$(6.1)\qquad \begin{aligned} & A(n,\nu;k;a,b;\sigma_1,\sigma_2) = \tau = \sum_h{}' \omega_a(h,k)\exp(-2\pi i(hn - h'\nu)/k) \\ & h \bmod k,\ h \equiv b\beta \pmod{p};\ \sigma_1 \leq h' < \sigma_2;\ 0 \leq \sigma_1 < \sigma_2 \leq k;\ p \mid k. \end{aligned}$$

Here Σ' means the sum over all integers prime to the modulus of the system, and h' denotes any solution of $hh' \equiv -1 \pmod{k}$. Examination of (2.81) shows $\omega_a(h,k)$ to be of period k. Because of the periodicity modulo k, we can change the modulus to Gk and select that h' which satisfies $hh' \equiv -1 \pmod{Gk}$. By inserting the value (4.9) for $\omega_a(h,k)$ and removing the restrictions on h and h', we find that $\tau = O(n^{1/3}k^{2/3+\epsilon})$. Hence, we secure the following result:[8]

---

[8] For details of this development, see [2], pages 649-651.

THEOREM 2. *The sum*

$$(6.2)\qquad \begin{aligned} &A(n,\nu;k;a,b;\sigma_1,\sigma_2)=\sum_h{}'\,\omega_a(h,k)\exp(-2\pi i(hn-h'\nu)/k)\\ &h \bmod k,\quad h\equiv b\beta \pmod p,\quad \sigma_1\leq h'<\sigma_2\end{aligned}$$

in which the parameters are all integers, $n>0$, $k>0$, $p\,|\,k$, $0\leq\sigma_1<\sigma_2\leq k$, *and* $a=1,2,\cdots,(p-1)/2$ *is subject to the estimate* $O(n^{1/3}k^{2/3+\epsilon})$ *uniformly in* ν, a, b, σ_1, σ_2.

By a similar method, and by use of the equations $G=pg$, $H=ph$, $HH'\equiv-1 \pmod k$, we obtain our next theorem.

THEOREM 3. *The sum*

$$(6.3)\qquad \begin{aligned} &B(n,\nu;k;\sigma_1,\sigma_2;a)=\sum_h{}'\,\chi_a(h,k)\exp(-2\pi i(hn-H'\nu)/k)\\ &h \bmod k,\quad HH'\equiv-1 \pmod k,\quad \sigma_1\leq h'<\sigma_2\end{aligned}$$

has the estimate $O(n^{1/3}k^{2/3+\epsilon})$ *uniformly in* ν, σ_1, σ_2, a.

III. A Convergent Series for $p_a(n)$.

7. We now apply the Hardy-Littlewood method, with the sharper estimate of Kloosterman and Rademacher,[9] to the generating functions $F_a(x)$. We obtain

$$p_a(n)=(1/2\pi i)\int_C (F_a(x)/x^{n+1})\,dx$$

$$(7.1)\qquad =\sum_{h,k}{}'\exp(-2\pi ihn/k)\int_{-\theta'}^{\theta''}F_a(\exp[2\pi ih/k-2\pi(z/k)])\exp(2\pi nw)\,d\phi,$$

$$0\leq h<k\leq N,\qquad \theta'=\theta'_{h,k},\qquad \theta''=\theta''_{h,k}.$$

We now apply the transformation formulas, and write

$$(7.21)\qquad p_a(n)=p_a^{(1)}(n)+p_a^{(2)}(n)$$

where $p_a^{(1)}(n)$ is the sum of all terms for which $p\,|\,k$ and $p_a^{(2)}(n)$ is the same for $p\nmid k$. Choose β such that $a\beta\equiv+1 \pmod p$. This together with (2.52) shows that

$$h\equiv b\beta \pmod p.$$

Then we obtain from (7.1) and the remarks above

---

[9] See [5]; the notations here are those of Rademacher.

$$\begin{aligned} p_a^{(1)}(n) &= \sum_{b=1}^{(p-1)/2} \sum_{h,k}{}' \omega_a(h,k) \exp\left(-2\pi i(hn/k)\right) \\ (7.22) \quad &\times \int_{-\theta'}^{\theta''} \sum_{\nu=0}^{\infty} p_b(\nu) \exp\left(2\pi i h'\nu/k\right) \exp\{-(\pi/k^2 w)\left(2\nu - (B/6p)\right) \\ &\qquad + \pi w\left(2n - (A/6p)\right)\} d\phi, \end{aligned}$$

$$0 \leq h < k \leq N, \qquad h \equiv b\beta \pmod{p}, \qquad \theta' = \theta'_{h,k}, \qquad \theta'' = \theta''_{h,k}.$$

The sum over ν is now split into two parts, $Q(n)$ and $R(n)$ depending upon B and *a fortiori* b such that in $Q(n)$ the coefficient of w^{-1} is positive and in $R(n)$ the coefficient of w^{-1} is negative. $R(n)$ contributes nothing to the final result; in fact,

$$(7.3) \qquad R(n) = O\left(\exp\left(c\pi N^{-2}\right) n^{1/3} N^{-1/3+\epsilon}\right).$$

Referring to (7.21) we note that $Q(n)$ contains all ν such that $\nu \leq [B/12p]$. If B is negative, the sum over ν is vacuous. For positive B and $\nu = 0, 1, \cdots, [B/12p]$

$$p_b(\nu) = \begin{cases} 0 \text{ if } b \nmid \nu \\ 1 \text{ if } b \mid \nu. \end{cases}$$

Here $p_b(\nu)$ denotes the number of partitions of ν into parts congruent to $\pm b \pmod{p}$ and since

$$\nu < B/12p = \tfrac{1}{2}\, \frac{p^2 - 6pb + 6b^2}{6p} = \tfrac{1}{2}(p/6 - b + b^2/p) < p/12$$

those parts that must be congruent to $\pm b \pmod{p}$ must be b itself, for $p - b$ or $p + b$ is too large $(p - b > p/2 > p/12 > \nu)$. Also, $B \geq 0$ means $b^2 - pb + p^2/6 \geq 0$, i.e., $b^2 - pb + p^2/4 \geq p^2/12$. Thus, the only permissible b for $Q(n)$ are $b = 1, 2, \cdots, [p(3 - \sqrt{3})/6]$. Setting $\nu = b\rho$, we find

$$\begin{aligned} Q(n) &= \sum_{b=1}^{[p(3-\sqrt{3})/6]} \sum_{h,k}{}' \omega_a(h,k) \exp\left(-2\pi i(hn/k)\right) \sum_{\rho=0}^{[B/12pb]} \exp\left(2\pi i h' b\rho/k\right) \\ (7.4) \quad &\times \int_{-\theta'}^{\theta''} \exp\{-(\pi/k^2 w)\left(2b\rho - (B/6p)\right) + \pi w\left(2n - (A/6p)\right)\} d\phi, \end{aligned}$$

$$0 \leq h < k \leq N, \qquad h \equiv b\beta \pmod{p}.$$

Evaluating $Q(n)$, we find

$$(7.5) \quad Q(n) = 2\pi \sum_{k=1}^{N} \sum_{b=1}^{[p(3-\sqrt{3})/6]} \sum_{\rho=0}^{[B/12pb]} A_{k,b}(n) L_k(n) + O\left(\exp\left(c\pi N^{-2}\right) n^{1/3} N^{-1/3+\epsilon}\right)$$

where

$$A_{k,b}(n) = A(n, \nu; k; a, b; 0, k) \tag{7.51}$$

and

$$\begin{aligned} L_k(n) &= \frac{(B-12pb\rho)^{1/2}}{k(12pn-A)^{1/2}} \; I_1\left\{\frac{\pi(12pn-A)^{1/2}(B-12pb\rho)^{1/2}}{3pk}\right\} \text{ for } n > A/12p \\ &= \frac{(B-12pb\rho)^{1/2}}{k(A-12pn)^{1/2}} \; J_1\left\{\frac{\pi(A-12pn)^{1/2}(B-12pb\rho)^{1/2}}{3pk}\right\} \text{ for } n < A/12p. \end{aligned} \tag{7.52}$$

Now

. 6)

$$\begin{gathered} p_a{}^{(2)}(n) = \tfrac{1}{2}\sum_{h,k}{}' \chi_a(h,k) \csc \pi\alpha/p \exp(-2\pi i(hn/k)) \\ \times \int_{-\theta'}^{\theta''} \sum_{\nu=0}^{\infty} c_a(\nu) \exp(2\pi i H'\nu/k) \exp\{-(\pi/Kkw)(2\nu-(1/6)) + \pi w(2n-(A/6p))\} d\phi, \\ 0 \leq h < k \leq N, \qquad \theta' = \theta'_{h,k}, \qquad \theta'' = \theta''_{h,k}. \end{gathered}$$

As before, we split this sum into two parts $Q'(n)$ and $R'(n)$. Here $Q'(n)$ contains only the value $\nu = 0$. We find, as before,

$$R'(n) = O(\exp(c\pi N^{-2}) n^{1/3} N^{-1/3+\epsilon}) \tag{7.7}$$

and

$$Q'(n) = \pi \sum_{k=1}^{N} \csc(\pi\alpha/p) B_k(n) L'_k(n) + O(\exp(c\pi N^{-2}) n^{1/3} N^{-1/3+\epsilon}). \tag{7.8}$$

where

$$B_k(n) = B(n, 0; k; 0, k; a) \tag{7.81}$$

and

$$\begin{aligned} L'_k(n) &= \frac{1}{k(12pn-A)^{1/2}} \; I_1\left\{\frac{\pi(12pn-A)^{1/2}}{3pk}\right\} \text{ for } n > A/12p \\ &= \frac{1}{k(A-12pn)^{1/2}} \; J_1\left\{\frac{\pi(A-12pn)^{1/2}}{3pk}\right\} \text{ for } n < A/12p. \end{aligned} \tag{7.82}$$

Upon combining (7.3), (7.5), (7.7), and (7.8), keeping n fixed and letting $N \to \infty$, we obtain our desired result.

THEOREM 4. *The number, $p_a(n)$, of partitions of a positive integer n into positive summands of the form $pl \pm a$ $(a = 1, 2, \cdots, (p-1)/2)$ is given by the convergent series*

$$\begin{aligned} p_a(n) &= 2\pi \sum_{\substack{k>0 \\ p\mid k}} \sum_{b=1}^{[p(3-\sqrt{3})/6]} \sum_{\rho=0}^{[B/12pb]} A_{k,b}(n) L_k(n) \\ &\quad + \pi \sum_{\substack{k>0 \\ p\nmid k}} B_k(n) \left| \csc \pi a k/p \right| L'_k(n) \end{aligned} \tag{7.9}$$

where $L_k(n)$ *and* $L'_k(n)$ *are given by* (7.52) *and* (7.82), $A_{k,b}(n)$ *by* (7.51), *and* $B_k(n)$ *by* (7.81).

8. To determine the asymptotic ratios of the $p_a(n)$, we have to investigate the dominant terms in each of the series of (7.9). Inspection of these series clearly shows that the leading term in each series dominates all others. Thus, we must have

$$\begin{aligned} &k = p, \quad b = 1, \quad \rho = 0 \quad \text{if} \quad p \mid k \\ &k = 1 \quad \text{if} \quad p \nmid k. \end{aligned} \tag{8.1}$$

In brief, the dominant terms are

$$\frac{2\pi}{(12pn - A)^{1/2}} A_{p,1}(n)\,(1/p) B_1^{1/2} \quad I_1 \left\{ \frac{\pi(12pn - A)^{1/2} B_1^{1/2}}{3p^2} \right\} \tag{8.2}$$

in the series for $p \mid k$ and

$$\frac{\pi}{(12pn - A)^{1/2}} B_1(n) \,|\csc \pi a/p\,| \quad I_1 \left\{ \frac{\pi(12pn - A)^{1/2}}{3p} \right\} \tag{8.3}$$

in the series for $p \nmid k$.

From (2.81) and (8.1), we find

$$B_1 = p^2 - 6p + 6.$$

Hence,

$$B_1/p^2 = \frac{p^2 - 6p + 6}{p^2} = 1 - 6/p + 6/p^2 < 1. \tag{8.4}$$

Therefore

$$\frac{\pi(12pn - A)^{1/2} B_1^{1/2}}{3p^2} < \frac{\pi(12pn - A)^{1/2}}{3p}\,. \tag{8.5}$$

From the theory of the Bessel functions, we obtain the following estimates[10]

$$\begin{aligned} I_1(z) &= O(z), & |z| < 1 \\ I_1(z) &\sim \exp z \cdot (2\pi z)^{-1/2}, & |z| > 1. \end{aligned}$$

By virtue of (8.2), (8.3), (8.4), and (8.5) we note that the term originating from the series for the case $p \nmid k$ dominates that originating from the series for the case $p \mid k$. Hence we obtain

---

[10] See [8], page 203.

$$p_1(n) : p_2(n) : \cdots : p_{(p-1)/2}(n)$$
$$\sim \frac{|\csc \pi/p|}{(12pn - A_1)^{1/2}} I_1 \left\{ \frac{\pi(12pn - A_1)^{1/2}}{3p} \right\} :$$
$$\frac{|\csc 2\pi/p|}{(12pn - A_2)^{1/2}} I_1 \left\{ \frac{\pi(12pn - A_2)^{1/2}}{3p} \right\} : \cdots :$$
$$\frac{|\csc (p-1)\pi/2p|}{(12pn - A_{(p-1)/2})^{1/2}} I_1 \left\{ \frac{\pi(12pn - A_{(p-1)/2})^{1/2}}{3p} \right\}.$$

Applying (8. 6), we finally obtain

$$p_1(n) : p_2(n) : \cdots : p_{(p-1)/2}(n)$$
$$\sim \csc \pi/p : \csc 2\pi/p : \cdots : \csc (p-1)\pi/2p.$$

UNIVERSITY OF PENNSYLVANIA.

BIBLIOGRAPHY.

1. M. Haberzetle, "On some partition functions," *American Journal of Mathematics*, vol. 63 (1941), pp. 589-599.
2. J. Lehner, "A partition function connected with the modulus five," *Duke Mathematical Journal*, vol. 8 (1941), pp. 631-655.
3. I. Niven, "On a certain partition function," *American Journal of Mathematics*, vol. 62 (1940), pp. 353-364.
4. H. Rademacher, "On the partition function $p(n)$," *Proceedings of the London Mathematical Society*, (2), vol. 43 (1937), pp. 241-254.
5. H. Rademacher, "The Fourier coefficients of the modular invariant $J(\tau)$," *American Journal of Mathematics*, vol. 60 (1938), pp. 501-512.
6. H. Rademacher and A. L. Whiteman, "Theorems on Dedekind sums," *American Journal of Mathematics*, vol. 63 (1941), pp. 377-407.
7. H. Salié, "Zur Abschätzung der Fourierkoeffizienten ganzer Modulformen," *Mathematische Zeitschrift*, vol. 36 (1933), pp. 263-278.
8. G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge, 1922.

ON THE LACK OF AN EUCLIDEAN ALGORITHM IN $R(\sqrt{61})$.*

By L. K. HUA and W. T. SHIH.

It has been proved by A. Brauer[1] that the fields $R(\sqrt{p})$, p a prime $\equiv 13 \pmod{24}$, are non-Euclidean, except possibly for $p = 13, 37, 61$ and 109. In the cases $p = 13$ and 37 the fields $R(\sqrt{p})$ are Euclidean. Thus it is an open question whether the fields $R(\sqrt{61})$ and $R(\sqrt{109})$ are Euclidean or not,

THEOREM 1. *The quadratic field $R(\sqrt{61})$ is non-Euclidean*

It is known that Theorem 1 is a consequence of

THEOREM 2. *For any integers x, y we have always*

$$|(x + (1/2)y - 1/78)^2 - (61/4)(y - 17/39)^2| > 1.$$

Let

$$u = 78x + 39y - 1, \qquad v = 39y - 17,$$

If Theorem 2 is false, then we have integers u, v and k such that

$$u^2 - 61v^2 = k, \text{ where } |k| \leq 78^2 = 6084. \tag{1}$$

Since

$$u \equiv -1, \qquad v \equiv -17 \pmod{39}, \tag{2}$$

we have

$$k \equiv 1^2 + 17^3 \equiv 0 \pmod{39};$$

furthermore since

$$k \equiv (y+1)^2 - (y+1)^2 = 0 \pmod{4},$$

we have that k is a multiple of 156 $(= 4 \times 39)$. Moreover, evidently $\left(\frac{k}{61}\right) = 1$, since k is not a multiple of 61 (in fact $61 \times 156 > 6084$). By these two properties we may list all the possible values of l, where $k = 156l$ and $|l| \leq 39$:

0, ± 1, ± 3, ± 4, ± 5, ± 9, ± 12, ± 13, ± 14, ± 15, ± 16, ± 19, ± 20, ± 22, ± 25, ± 27, ± 34, ± 36, ± 39.

---

* Received April 25, 1944.

[1] Brauer, *American Journal of Mathematics*, vol. 62 (1940), pp. 697-716.

For $l = 0, \pm 14, \pm 22, \pm 34$, (1) is insoluble. In fact, e. g., if

$$u^2 - 61v^2 = \pm 14 \times 156,$$

then

$$u^2 \equiv 61v^2 \pmod 7,$$

and $\left(\frac{61}{7}\right) = -1$, which is impossible.

We now give the solutions of (1) for these possible values of k.

LEMMA 1. *The numbers of primary*[2] *solutions of* (1) *are given by the following table:*

l	± 1	± 3	± 4	± 5	± 9	± 12	± 13	± 15	± 16	± 19	± 20	± 25	± 27	± 36	± 39
no. of primary solutions of (1)	4	6	4	8	8	6	6	12	4	8	8	12	10	8	9

The proof of the lemma is a straightforward calculation of

$$\sum_{n \mid k} \left(\frac{61}{n}\right),$$

which is the number of primary representations of k.

LEMMA 2. *If* $u^2 - 61v^2 = k$ *has no primary solution satisfying*

$$(2') \qquad u = \pm 1, \qquad v = \pm 17 \pmod{39},$$

then $u^2 - 61v^2 = \pm k$ *has no solution satisfying* (2′).

Proof. Since $\eta = (39 + 5\sqrt{61})/2$ is the fundamental solution of $u^2 - 61v^2 = -4$ (in fact $39^2 - 61 5^2 = -4$), the solutions of $u^2 - 61v^2 = \pm k$ are given by

$$\frac{u + v\sqrt{61}}{2} = \pm \left(\frac{39 + 5\sqrt{61}}{2}\right)^{\lambda} \left(\frac{u_0 + v_0\sqrt{61}}{2}\right), \qquad \lambda \lesseqgtr 0,$$

where u_0, v_0 is a solution of (1).

For $\lambda = 1$, we have

$$(3) \qquad u = \pm \tfrac{1}{2}(39u_0 + 5 \times 61v_0), \qquad v = \pm \tfrac{1}{2}(39v_0 + 5u_0),$$

i. e.

$$u \equiv \pm 16v_0, \qquad v \equiv \pm 17u_0, \qquad \pmod{39},$$

i. e.

$$u_0 \equiv \pm 16v, \qquad v_0 \equiv \pm 17u, \qquad \pmod{39}$$

(since $16 \times 17 \equiv -1 \pmod{39}$). If $u \equiv \pm 1$, $v \equiv \pm 17$, then we have $u_0 \equiv \pm 1$, $v_0 \equiv \pm 17 \pmod{39}$.

---

[2] Landau, *Vorlesungen über Zahlentheorie*, Bd. 1, S. 140.

The proof is easily completed by repeated application of the same argument.

Thus in order to prove Theorem 2 it remains only to establish that there is no primary solution of (1) with (2) and $0 < l \leq 39$.

LEMMA 3. *All the primary solutions of* (1) *are given by the following table*:

l	u, v
1	20, 2; 41, 5; 1484, 190; 3335, 427.
3	23, 1; 38, 4; 282, 36; 633, 81; 5147, 659; 11567, 1481.
4	40, 4; 82, 10; 2968, 380; 6670, 854.
5	29, 1; 32, 2; 151, 19; 337, 43; 883, 113; 1984, 254; 12473, 1597; 16136, 2066.
9	60, 6; 123, 15; 245, 31; 548, 70; 977, 125; 2195, 281; 4452, 570; 10005, 1281.
12	46, 2; 76, 8; 564, 72; 1266, 162; 10294, 1318; 23134, 2962.
13	97, 11; 208, 26; 463, 59; 1672, 214; 3757, 481; 8443, 1081.
15	49, 1; 73, 7; 134, 16; 171, 21; 293, 37; 378, 48; 2367, 303; 3062, 392; 5319, 681; 6881, 881; 13957, 1787; 31366, 4016.
16	80, 8; 164, 20; 5936, 760; 13340, 1708.
19	55, 1; 67, 5; 433, 55; 970, 124; 1165, 149; 2617, 335; 21283, 2725; 35935, 4601.
20	58, 2; 64, 4; 302, 38; 674, 86; 1766, 226; 3968, 508; 24946, 3194; 32272, 4132.
25	83, 7; 100, 10; 161, 19; 205, 25; 527, 67; 1181, 151; 1259, 161; 2828, 362; 7420, 950; 9599, 1229; 16675, 2135; 21572, 2762.
27	69, 3; 114, 12; 191, 23; 419, 53; 846, 108; 1899, 243; 3851, 493; 8654, 1108; 15441, 1977; 34701, 4443.
36	120, 12; 246, 30; 490, 62; 1096, 140; 1954, 250; 4390, 562; 8904, 1140; 20010, 2562.
39	78, 0; 105, 9; 322, 40; 715, 91; 1447, 185; 1603, 205; 3250, 416; 7303, 935; 26430, 3384.

By Lemma 1, the table contains all primary solutions. The verification is straightforward.

Moreover no solution contained in the table satisfies (2′). Thus Theorem 2 is established.

NATIONAL TSING HUA UNIVERSITY.

THE PROBLEM OF MILLOUX FOR FUNCTIONS ANALYTIC THROUGHOUT THE INTERIOR OF THE UNIT CIRCLE.*

By Maurice H. Heins.

---

1. Introduction. Let E denote a point set of the z-plane lying in, and closed relative to, $|z| < 1$, having the property that for all R satisfying: $0 \leqq R < 1$ the intersection of E with $|z| = R$ is not void. A function $f(z)$ will be said *to belong to the class* $\mathfrak{F}_m$, if $f(z)$ is defined, analytic, and of modulus less than one for $|z| < 1$ and if there exists an E such that $z \in E$ implies $|f(z)| \leqq m$, where m is a given positive number less than one. Let $M(f; r)$ have its usual significance as $\max_{|z|=r} |f(z)|$. *It is required to determine* $\text{l.u.b.}_{f \in \mathfrak{F}_m} M(f; r)$ $(0 < r < 1)$ *and the associated extremal functions.*

This problem is solved completely in the present paper; it will be shown that the extremal function suitably normalized is unique and independent of r. The normalized extremal function will be determined and expressed with the aid of the Jacobi theta functions. Two methods of obtaining the extremal function are indicated. One involves the theory of Blaschke products; the other gives an algorithm for determining the extremal function by the repeated composition of $(1, 2)$ directly conformal maps of the interior of the unit circle onto itself.

The methods developed are applicable to other problems. Analogues of Tchebycheff polynomials in the hyperbolic plane are defined and determined.

The problem of the present paper is closely related to a series of investigations initiated by Milloux [9], who showed that, if $f(z)$ is analytic and of modulus less than or equal to one for $|z| \leqq 1$ and if there exists a Jordan arc J lying in $|z| < 1$, except for one endpoint on $|z| = 1$, and joining $z = 0$ to the circumference $|z| = 1$ with $|f(z)| \leqq m$ (< 1) for $z \in J$, then there exists a positive constant k, *independent of m and J*, such that

$$(1.1) \qquad |f(z)| \leqq m^{k[1-|z|]} \qquad (|z| < 1).$$

Lower and upper estimates for the best value of k have been given by Landau [8]. Subsequently, E. Schmidt [14] considered the problem of Milloux under the weaker hypotheses which assumed: (a) $f(z)$ to be analytic and of modulus less than one for $z \in \mathfrak{G}$ where $\mathfrak{G} \equiv [|z| < 1] \cdot CJ$, J having the same connota-

---

* Received January 11, 1944.

tion as above; (b) $\overline{\lim}_{z(\in \mathcal{G})\to J} |f(z)| \leqq m$, and showed the existence of a positive k, again independent of m and J, such that (1.1) held for the $f(z)$ considered and $z \in \mathcal{G}$. In fact, he showed the best value of k for the problem he treated to be $1/\pi$ and indicated that this was also the best value of k when $f(z)$ is actually defined and analytic throughout $|z| < 1$. The problem was later studied under broader hypotheses by R. Nevanlinna [12], Beurling [2], Fenchel [4], and Milloux [10], but always with the view to obtaining inequalities having universal constants independent of m and the set E where $\overline{\lim}_{z\to E} |f(z)| \leqq m$.

It should be emphasized that the present investigation is concerned with functions analytic throughout the interior of the unit circle and deals with m fixed.

2. Preliminary reductions. Let $g(z)$ be defined, analytic and of modulus less than one for $|z| < 1$. Recall that $g(z)$ can be represented[1] by

$$g(z) \equiv \pi(z) \exp \left\{ \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \, d\mu(\theta) + i\lambda \right\}, \tag{2.1}$$

where $\mu(\theta)$ is defined and monotone non-increasing for $0 \leqq \theta \leqq 2\pi$, λ is a real constant and $\pi(z)$ is defined to be (a) identically equal to one if $g(z)$ has no zeros for $|z| < 1$, (b) the finite product

$$\prod_{k=1}^{n} \frac{\bar{\alpha}_k}{|\alpha_k|} \cdot \frac{\alpha_k - z}{1 - \bar{\alpha}_k z} \tag{2.2}$$

if $g(z)$ has a finite number of zeros $\alpha_1, \cdots, \alpha_n$, multiplicities being counted as usual (if any $\alpha_k = 0$, then the corresponding factor is to be replaced simply by z), (c) the Blaschke product[2]

$$\prod_{k=1}^{\infty} \frac{\bar{\alpha}_k}{|\alpha_k|} \cdot \frac{\alpha_k - z}{1 - \bar{\alpha}_k z} \tag{2.3}$$

if $g(z)$ has an infinite set of zeros $\alpha_1, \alpha_2, \cdots$, the same conventions prevailing here and throughout the discussion as in the case of the finite product (2.2).

Suppose now that $f_0(z) \in \mathcal{F}_m$ is extremal and that E_0 is a corresponding E-set.[3] By means of a trivial normalization it may be assumed that $f_0(r)$

---

[1] This is essentially the representation due to Herglotz [6]. Cf. Evans [3] and R. Nevanlinna [11] for accounts of this representation. The use of the Poisson-Stieltjes representation is not necessary in this connection but is suggestive from a physical point of view.

[2] An exposition of the theory of Blaschke products as well as of the products (2.2) is to be found in Julia [7].

[3] The proof of the existence of an extremal function and an associated E-set is readily established and will be omitted.

$= M(f_0; r)$. The object of the present section is to show that, subject to this normalization, the set $[-1 < x \leqq 0]$ may be regarded as an associated E-set of $f_0(z)$. Define $\pi^*(z)$ to be identically equal to one if $f_0(z)$ is free from zeros;

$$\pi^*(z) \equiv \prod_{k=1}^{n} \frac{|z_k| + z}{1 + |z_k| z} \tag{2.3}$$

if $f_0(z)$ has a finite number of zeros $z_1, \cdots, z_n$;

$$\pi^*(z) \equiv \prod_{k=1}^{\infty} \frac{|z_k| + z}{1 + |z_k| z} \tag{2.4}$$

if $f_0(z)$ has an infinite set of zeros $z_1, z_2, \cdots$. Now let $h(z)$ denote

$$h(z) \equiv \pi^*(z) \exp \left\{ \frac{1-z}{1+z} \cdot \frac{1}{2\pi} \int_0^{2\pi} d\mu(\theta) \right\} \tag{2.5}$$

where $\mu(\theta)$ has the connotation of (2.1) for $f_0(z)$. If $z \in E_0$, then

$$|h(-|z|)| \leqq |f_0(z)|, \tag{2.6}$$

and

$$h(r) \geqq f_0(r), \tag{2.7}$$

equality prevailing in (2.7) if and only if $h(z) \equiv f_0(z)$. These facts may be readily ascertained by the use of standard inequalities for expressions of the form

$$\frac{z - \alpha}{1 - \bar{\alpha} z} \quad (|\alpha| < 1)$$

and for the Poisson-Stieltjes integral (inequalities of the Harnack type). Inequalities (2.6) and (2.7) imply, respectively, that $h(z) \in \mathfrak{F}_m$ with $[-1 < x \leqq 0]$ as an associated E-set and that $h(z)$ is extremal. Since the strong inequality prevails in (2.7) if $h(z) \not\equiv f_0(z)$, it follows that, if $f_0(z)$ is extremal, then $h(z) \equiv f_0(z)$, and hence

THEOREM 2.1. *If* $f_0(z)$ *is extremal and is so normalized that* $M(f_0; r) = f_0(r)$, *then* $[-1 < x \leqq 0]$ *is an associated E-set.*

3. The qualitative characterization of the extremal function. If $f_0(z)$ is extremal and normalized as in 2, then $f_0(z)$ is also an extremal function for the following problem.

PROBLEM. *Let* $f(z)$ *be analytic and of modulus less than one for* $|z| < 1$. *Further let it be required that* $|f(z)| \leqq m < 1$, *m being a given positive* *number less than one, for* z *real and non-positive* $(-1 < z \leqq 0)$. *Finally let* r *be real and lie strictly between zero and one. It is required to determine*

$$\underset{\{f\}}{\text{l. u. b.}}\,|f(r)|$$

and the associated extremal functions.

This problem has been studied qualitatively by the author,[4] the pertinent results being summarized in

THEOREM 3.1. *If* $f(z)$ *is so normalized that* $f(r) > 0$, *then there exists an extremal function,* $\phi(z; m, r)$, *and it is unique. It has the following properties:*

(a) $\phi(z; m, r)$ *is a Blaschke product;*

(b) $\phi(z; m, r)$ *has an infinite number of zeros, all simple and negative;*

(c) $\phi(z; m, r)$ *is real for* z *real;*

(d) *on the real axis between successive zeros of* $\phi(z; m, r)$ *there exists a unique point where* $\phi^2 = m^2$;

(e) $\phi(0; m, r) = m$, $\phi'(0; m, r) > 0$;

(f) ϕ' *has simple zeros at the points whose existence is affirmed in* (d) *and has no other zeros;*

(g) *the zeros of* $\phi^2 - m^2$ *are* $z = 0$ *and the points defined in* (d), *and there are no others.*

Let it be agreed that $G[w(z)]$ will denote the Riemannian image of $|z| < 1$, with respect to $w(z)$, where $w(z)$ is understood to be analytic for $|z| < 1$.

From Theorem 3.1 the structure of $G[\phi]$ can be readily inferred. It follows from (f) that $G[\phi]$ has branch-points only over $-m$ and m. Consider now the Riemann domain S lying over $|w| < 1$ which is defined as follows. Let S_1 denote the open simple disk $|w| < 1$ slit along the real axis from -1 to $-m$, and let S_k $(k \geqq 2)$ denote copies of the open disk $|w| < 1$ slit along the real axis from -1 to $-m$ and from m to $+1$. The Riemann domain S is constructed by joining the upper bank of the slit of S_1 to the lower bank of the left-hand slit of S_2 and the lower bank of the slit of S_1 to the upper bank of the left-hand slit of S_2, then by joining in the same fashion the banks of the right-hand slits of S_2 and S_3, then the banks of the left-hand slits of S_3 and S_4, and so forth. The Riemann domain, S, obtained in this fashion is simply-connected and bounded. Hence by the fundamental mapping theorem of conformal mapping there exists a function $w = \Phi(z; m)$ which is analytic and of modulus less than one for $|z| < 1$, which satisfies the normalizing conditions:

---

[4] Cf. M. H. Heins [5].

$$\Phi(0; m) = m, \qquad \Phi'(0; m) > 0,$$

and maps $|z| < 1$ onto the Riemann domain S.

Consider now the relation

$$\Phi[\psi(z); m] = \phi(z; m, r) \tag{3.1}$$

where $\psi(z)$ is normalized to vanish at $z = 0$. It is readily inferred from the structure of $S \equiv G[\Phi]$ that there exists a function, $\psi(z)$, which is analytic, single-valued, and of modulus less than one for $|z| < 1$, which vanishes at $z = 0$, and satisfies (3.1) identically for all z in $|z| < 1$. Next, from the structure of S and the normalization requirements imposed on Φ, it may be inferred that Φ is real and monotone increasing for z real and positive, and that Φ is real and of modulus not exceeding m for z real and negative. Finally, $\psi(z)$ is real and positive (in fact monotone increasing) for z real and positive. Hence $\Phi(z; m)$ is a competitor in the extremal problem proposed at the beginning of this section. Also $\psi(r) \leqq r$ $(0 < r < 1)$ where equality prevails if and only if $\psi(z) \equiv z$. From (3.1) and the enumerated properties of Φ and ψ it follows that

$$\Phi(r; m) \leqq \phi(r; m, r) = \Phi(\psi(r); m) \leqq \Phi(r; m). \tag{3.2}$$

Hence $\psi(r) = r$, $\Phi(z; m) \equiv \phi(z; m, r)$, and $S \equiv G[\phi]$. One important consequence of this result is

THEOREM 3.2. *The extremal function $\phi(z; m, r)$ is independent of r.*

For the remainder of the present paper the extremal function will be denoted simply by $\phi(z; m)$.

The information obtained about the structure of $G[\phi]$ will be very valuable for the explicit determination of $\phi(z; m)$. An immediate application of the present results is a recurrence relation for $\phi(z; m)$ which may be found as follows. Consider the function $\phi^*(z)$ defined by

$$[\phi^*(z)]^2 \equiv \frac{\phi(z; m) + m}{1 + m\phi(z; m)}; \qquad \phi^*(0) > 0. \tag{3.3}$$

The function $\phi^*(z)$ may be continued to be analytic throughout $|z| < 1$, since at the points where $\phi(z; m) = -m$ it is true that $\phi'(z; m)$ has a simple zero. The Riemann domain, $G[\phi^*]$, has the same ramification properties as $G[\phi]$ except that the branch points are now over $\sqrt{2m/(1+m^2)}$ and $-\sqrt{2m/(1+m^2)}$. The conditions (3.3) imposed on $\phi^*(z)$ imply that the derivative of $\phi^*(z)$ at $z=0$ is positive. Hence $\phi^*(z) \equiv \phi(z; \sqrt{2m/(1+m^2)})$ and (3.3) yields

$$[\phi(z; \sqrt{2m/(1+m^2)})]^2 \equiv \frac{\phi(z;m)+m}{1+m\phi(z;m)}. \tag{3.4}$$

4. The determination of $\varphi(z;m)$ by a process of successive compositions of rational functions. In this section an algorithm will be developed for determining $\phi(z;m)$ as the limit of a sequence of rational functions. One noteworthy feature of this algorithm is that each member of the sequence provides a majorant for $\phi(r;m)$ $(0<r<1)$ and hence furnishes incidentally an upper bound for $M(f;r)$ where $f \in \mathfrak{F}_m$.

Instead of considering $\phi(z;m)$ directly, it will be convenient to study

$$\omega(z) \equiv \frac{\phi(z;m)-m}{1-m\phi(z;m)}. \tag{4.1}$$

The structure of the Riemann domain, $G[\omega]$, is the same as that of $G[\phi]$ save that the ramification points of $G[\omega]$ lie over 0 and $-2m/(1+m^2)$. Let μ denote $2m/(1+m^2)$. Define $\omega_1(z)$ by the relation

$$\left[\frac{\omega_1(z)+k_1}{1+k_1\omega_1(z)}\right]^2 = \frac{\omega(z)+\mu}{1+\mu\omega(z)}, \tag{4.2}$$

and the requirements: $\omega_1(0)=0$, $\omega'_1(0)>0$, k_1 being real and so determined that these conditions are fulfilled.[5] It is clear that, subject to these conditions, $k_1{}^2=\mu$ and on differentiating (4.2) with respect to z that

$$2\,\frac{\omega_1(0)+k_1}{1+k_1\omega_1(0)}\cdot\frac{(1-k_1{}^2)\omega'_1(0)}{[1+k_1\omega_1(0)]^2} = \frac{(1-\mu^2)\omega'(0)}{[1+\mu\omega(0)]^2}$$

or

$$2k_1(1-k_1{}^2)\omega'_1(0) = (1-\mu^2)\omega'(0).$$

Hence $k_1>0$ and is equal to $\sqrt{\mu}$. With k_1 so specified, $\omega_1(z)$ is uniquely determined as a function analytic and of modulus less than unity for $|z|<1$. A study of the Riemann domain, $G[\omega_1]$, yields the result that $G[\omega_1]$ has the same structure as $G[\omega]$ but with its ramification points lying over 0 and $-(2\sqrt{\mu}/(1+\mu))$; $(2\sqrt{\mu}/(1+\mu)$ will be denoted by $\mu_1)$. From (4.2) it follows that $\omega_1(z)$ is real, monotone increasing and positive for z real and positive since $\omega(z)$ itself has this property. But then from (4.1) and (4.2)

---

[5] It is to be observed that this is essentially the device used by Koebe for establishing the fundamental mapping theorem.

$$(4.3)\quad \phi(z;m) \equiv \frac{\omega(z)+m}{1+m\omega(z)} = \frac{\left(\dfrac{\omega_1+\sqrt{\mu}}{1+\sqrt{\mu}\,\omega_1}\right)^2 + \left(\dfrac{m-\mu}{1-m\mu}\right)}{1+\left(\dfrac{m-\mu}{1-m\mu}\right)\left(\dfrac{\omega_1+\sqrt{\mu}}{1+\sqrt{\mu}\,\omega_1}\right)^2}$$

$$= \frac{\left(\dfrac{\omega_1+\sqrt{\mu}}{1+\sqrt{\mu}\,\omega_1}\right)^2 - m}{1-m\left(\dfrac{\omega_1+\sqrt{\mu}}{1+\sqrt{\mu}\,\omega_1}\right)^2}$$

and since, by Schwarz's lemma, $\omega_1(r) < r$ for r real and positive, a rational majorant for $\phi(r;m)$ is given by replacing $\omega_1(r)$ by r. Hence

$$(4.4)\qquad \phi(r;m) < \frac{\left(\dfrac{r+\sqrt{\mu}}{1+\sqrt{\mu}\,r}\right)^2 - m}{1-m\left(\dfrac{r+\sqrt{\mu}}{1+\sqrt{\mu}\,r}\right)^2}.$$

The process applied to $\omega(z)$ to determine $\omega_1(z)$ may now be applied to $\omega_1(z)$ with μ_1, $\sqrt{\mu_1}$, $\omega_1(z)$, $\omega_2(z)$ replacing μ, k_1, $\omega(z)$, $\omega_1(z)$, respectively, in (4.2) and the conditions $\omega_2(0)=0$, $\omega'_2(0)>0$ replacing the analogous conditions on $\omega_1(z)$. This process can be continued step-by-step indefinitely yielding a sequence of functions $\{\omega_k(z)\}$ and a sequence of positive constants $\{\mu_k\}$ $(k=0,1,\cdots)$ where

$$(4.5)\qquad \mu_0=\mu,\qquad \mu_{k+1}=\frac{2\sqrt{\mu_k}}{1+\mu_k}\qquad (k=0,1,\cdots);$$

$$(4.6)\qquad \omega_0(z)\equiv\omega(z),\ \left[\frac{\omega_{k+1}+\sqrt{\mu_k}}{1+\sqrt{\mu_k}\,\omega_{k+1}}\right]^2 \equiv \frac{\omega_k+\mu_k}{1+\mu_k\omega_k}\qquad (k=0,1,\cdots);$$

and

$$(4.7)\qquad \omega_k(0)=0,\qquad \omega'_k(0)>0\qquad (k=0,1,\cdots).$$

Let $q_k(Z)$ be defined by

$$(4.8)\qquad q_k(Z) \equiv \frac{\left(\dfrac{Z+\sqrt{\mu_k}}{1+\sqrt{\mu_k}\,Z}\right)^2 - \mu_k}{1-\mu_k\left(\dfrac{Z+\sqrt{\mu_k}}{1+\sqrt{\mu_k}\,Z}\right)^2}\qquad (k=0,1,\cdots)$$

and let $Q_k(Z)$ be defined by the relations

$$(4.9)\qquad Q_0(Z)\equiv q_0(Z);\qquad Q_k(Z)\equiv Q_{k-1}[q_k(Z)]\qquad (k=1,2,\cdots).$$

It is an immediate consequence of these definitions that

$$(4.10)\qquad \phi(z;m) \equiv \frac{Q_k[\omega_{k+1}(z)] + m}{1 + mQ_k[\omega_{k+1}(z)]} \qquad (k = 0, 1, \cdots).$$

An argument of the type used to deduce (4.4) yields the following sequence of majorants for $\phi(r:m)$ (r real and between 0 and 1):

$$(4.11)\qquad \phi(r;m) < \frac{Q_k(r) + m}{1 + mQ_k(r)} \qquad (k = 0, 1, \cdots).$$

It is natural to conjecture that

$$\phi(z;m) = \lim_{k\to\infty} \frac{Q_k(z) + m}{1 + mQ_k(z)} \quad \text{for} \quad |z| < 1.$$

This may be proved from properties of $Q_k(z)$ resulting solely from its definition. From (4.9) it follows that the degree of $Q_k(z)$ is 2^{k+1}. Second, consider the image of the interval $[-\mu_{k+1} \leqq x \leqq 0]$ with respect to $Q_k(x)$. For $k = 0$, the image of $[-\mu_1 \leqq x \leqq 0]$ with respect to $Q_0(x)$ is $[-\mu \leqq u \leqq 0]$. This may be seen by observing that the minimum value attained by $Q_0(x)$ for real x is $-\mu$ and that this minimum is attained at $x = -\sqrt{\mu}$ which evidently satisfies the inequality

$$-\mu_1 = -\frac{2\sqrt{\mu}}{1+\mu} < -\sqrt{\mu} < 0.$$

To obtain the image of $[-\mu_2 \leqq x \leqq 0]$ with respect to $Q_1(x)$, recall that $Q_1(z) \equiv Q_0[q_1(z)]$ and observe that the image of $[-\mu_2 \leqq x \leqq 0]$ with respect to $q_1(x)$ is $[-\mu_1 \leqq X \leqq 0]$; this last remark follows by the argument used on $Q_0(x)$ and $[-\mu_1 \leqq x \leqq 0]$. Hence the image of $[-\mu_2 \leqq x \leqq 0]$ with respect to $Q_1(x)$ is $[-\mu \leqq u \leqq 0]$. By observing that the image of $[-\mu_{k+1} \leqq x \leqq 0]$ with respect to $q_k(x)$ is $[-\mu_k \leqq X \leqq 0]$ and recalling the defining relations (4.9) the proof of the original assertion is established.

To complete the proof of the conjecture, note that $\lim_{k\to\infty} \mu_k = 1$ (this fact follows from the elementary observation that $\{\mu_k\}$ is a positive monotone increasing sequence with limit $\mu^* \leqq 1$, and that $\mu^* = 2\sqrt{\mu^*}/(1 + \mu^*)$). Let $\{Q_{k(n)}(z)\}$ denote any subsequence of $\{Q_k(z)\}$ which converges for $|z| < 1$ and let $Q^*(z)$ denote the corresponding limit function. Since for each k ($k = 0, 1, 2, \cdots$) the image of $[-\mu_{k+1} \leqq x \leqq 0]$ with respect to $Q_k(z)$ is $[-\mu \leqq u \leqq 0]$ and since $\lim_{k\to\infty} \mu_k = 1$, the limit function $Q^*(z)$ must have the property that the image of $[-1 < x \leqq 0]$ with respect to $Q^*(z)$ must belong to the interval $[-\mu \leqq u \leqq 0]$. Hence the function

$$\frac{Q^*(z)+m}{1+mQ^*(z)}$$

must be a competitor in the extremal problem of **3**. But then it is clear that

$$\frac{Q^*(r)+m}{1+mQ^*(r)} \leqq \phi(r;m),$$

and on the other hand (4.11) implies

$$\phi(r;m) \leqq \frac{Q^*(r)+m}{1+mQ^*(r)}.$$

In other words,

$$\phi(z;m) \equiv \frac{Q^*(z)+m}{1+mQ^*(z)}.$$

Now the sequence $\{Q_k(z)\}$ does converge. If it did not, there would exist two subsequences with distinct limit functions. However the argument just applied to $Q^*(z)$ shows that any limit function must coincide with

$$\frac{\phi(z;m)-m}{1-m\phi(z;m)}.$$

Hence $\lim_{k\to\infty} Q_k(z)$ exists for $|z|<1$ and stands in the indicated relation to $\phi(z;m)$.

THEOREM 4.1. *If* $f(z) \in \mathfrak{F}_m$, *then* $M(f;r) < \dfrac{Q_k(r)+m}{1+mQ_k(r)}$ $(k=0,1,2,\cdots)$. *Further, the sequence* $\{Q_k(z)\}$ *converges continuously for* $|z|<1$ *and*

$$\phi(z;m) \equiv \frac{\lim_{k\to\infty} Q_k(z)+m}{1+m[\lim_{k\to\infty} Q_k(z)]}.$$

5. The representation of $\varphi(z;m)$ as a Blaschke product. Let $B(z;m)$ denote $\phi(-z^2;m)$. Clearly $B(z;m)$ is a Blaschke product which is even and has all its zeros simple and real. Another consequence of the definition of $B(z;m)$ is that $|B(z;m)|$ has m as its maximum between successive zeros of $B(z;m)$ (for z real). The Riemann domain, $G[B(z;m)]$, is completely ramified (the covering is always twofold locally) over m and $-m$. Hence if α denotes the smallest positive zero of $B(z;m)$, and if T denotes the linear fractional transformation of $|z|<1$ onto itself with fixed points -1 and $+1$, which carries $-\alpha$ into α, then

$$B(T;m) = -B(z;m). \tag{5.1}$$

The zeros of $B(z;m)$ are therefore $\{T^k\alpha\}$ $(k=0,1,2,\cdots;-1,-2,\cdots)$.

For by (5.1) $T^k\alpha$ is a zero of $B(z;m)$ for every integer k. Nor can there be any other zero since such a zero would have a homologue with respect to some integral power of T strictly in the open interval $(-\alpha, \alpha)$ and this is contrary to the definition of α. It now remains to determine α and T in terms of m.

The transformation T may be expressed in the form

$$\frac{Tz-1}{Tz+1} = \lambda\,\frac{z-1}{z+1} \qquad (\lambda \text{ real, } 0 < \lambda < 1). \tag{5.2}$$

The condition, $T(-\alpha) = \alpha$, implies that $\lambda = \left(\frac{1-\alpha}{1+\alpha}\right)^2$. Hence T^k is given by

$$\frac{T^kz-1}{T^kz+1} = \left(\frac{1-\alpha}{1+\alpha}\right)^{2k}\cdot\ \frac{z-1}{z+1}, \tag{5.3}$$

or

$$T^kz = \frac{1-\left(\frac{1-\alpha}{1+\alpha}\right)^{2k}\left(\frac{1-z}{1+z}\right)}{1+\left(\frac{1-\alpha}{1+\alpha}\right)^{2k}\left(\frac{1-z}{1+z}\right)}. \tag{5.4}$$

Since $B(z;m)$ is an even Blaschke product which is positive for $z=0$, it must necessarily be given by

$$B(z;m) = \prod_{k=0}^{\infty} \frac{(T^k\alpha)^2 - z^2}{1-(T^k\alpha)^2z^2}. \tag{5.5}$$

The value of α is determined by the condition $B(0;m) = m$. In other words, the relation

$$m = \prod_{k=0}^{\infty} (T^k\alpha)^2 = \prod_{k=0}^{\infty} \left[\frac{1-\left(\frac{1-\alpha}{1+\alpha}\right)^{2k+1}}{1+\left(\frac{1-\alpha}{1+\alpha}\right)^{2k+1}}\right]^2 \tag{5.6}$$

must hold. It is readily verified that there exists a unique real α lying between 0 and 1 which satisfies (5.6) for a given positive m less than 1. This may be deduced by observing that the linear fractional transformation $\alpha \,\Big|\, \frac{1-\alpha}{1+\alpha}$ carries the interval $(0 < \alpha < 1)$ into itself and that the infinite product

$$P(\tau) \equiv \prod_{k=0}^{\infty} \left[\frac{1-\tau^{2k+1}}{1+\tau^{2k+1}}\right]^2$$

is analytic for $|\tau| < 1$ and defines a monotone strictly decreasing map of the interval $(0 < \tau < 1)$ onto itself. As a matter of fact, the inversion of $P(\tau)$ is a classical problem in the theory of theta functions.

To recast the present problem in terms of the theory of theta functions, recall the following definitions and identities.[6] The functions $\theta_4(u)$ and $\theta_3(u)$ with parameter τ ($|\tau| < 1$) can be defined by

$$\begin{aligned} \theta_4(u) &\equiv 1 - 2\tau\cos 2\pi u + 2\tau^4\cos 4\pi u - 2\tau^9\cos 6\pi u + \cdots; \\ \theta_3(u) &\equiv 1 + 2\tau\cos 2\pi u + 2\tau^4\cos 4\pi u + 2\tau^9\cos 6\pi u + \cdots. \end{aligned} \tag{5.7}$$

For each τ satisfying $|\tau| < 1$ the functions $\theta_4(u)$ and $\theta_3(u)$ are entire in u. For the present purposes the following identities are significant:

$$\begin{aligned} \theta_4(u) &\equiv C\prod_{k=0}^{\infty}(1 - 2\tau^{2k+1}\cos 2\pi u + \tau^{4k+2}); \\ \theta_3(u) &\equiv C\prod_{k=0}^{\infty}(1 + 2\tau^{2k+1}\cos 2\pi u + \tau^{4k+2}), \end{aligned} \tag{5.8}$$

where C is independent of z and is obviously not zero—its exact value is of no concern for the problem at hand—and

$$\begin{aligned} \theta_4(0) &= C\prod_{k=0}^{\infty}(1 - \tau^{2k+1})^2; \\ \theta_3(0) &= C\prod_{k=0}^{\infty}(1 + \tau^{2k+1})^2. \end{aligned} \tag{5.9}$$

From (5.6) and (5.9)

$$m = \theta_4(0)/\theta_3(0) \tag{5.10}$$

where the parameter τ is equal to $(1-\alpha)/(1+\alpha)$. The formulas (5.8) will be employed in **6** to represent $\phi(z;m)$ in terms of the functions $\theta_4(u)$ and $\theta_3(u)$.

With α assumed determined as a function of m, $\phi(z;m)$ may be expressed immediately as a Blaschke product. The definition of $B(z;m)$ and the formula (5.5) imply

$$\phi(z;m) \equiv \prod_{k=0}^{\infty} \frac{(T^k\alpha)^2 + z}{1 + (T^k\alpha)^2 z}. \tag{5.11}$$

Note that the sequence $\{\phi_N(z)\}$, where $\phi_N(z) \equiv \prod_{k=0}^{N} \frac{(T^k\alpha)^2 + z}{1 + (T^k\alpha)^2 z}$, is monotone decreasing for any z real, positive, and less than 1. Hence $\phi_N(r)$ furnishes a rational majorant for $M(f;r)$ where $f \in \mathfrak{F}_m$ for $N = 0, 1, 2, \cdots$. By way of comparison with the methods developed in **4**, it is to be observed that the determination of the coefficients of ϕ_N involves the transcendental operation of computing α in terms of m, whereas the coefficients of $Q_k(z)$ may be found by essentially "algebraic" operations.

---

[6] Following the notation used by Whittaker and Watson [16].

6. The representation of $\varphi(z; m)$ in terms of theta functions. With $\tau = (1 - \alpha)/(1 + \alpha)$ formulas (5.4) and (5.11) imply

$$\phi(z; m) = \prod_{k=0}^{\infty} \frac{\left(\dfrac{1 - \tau^{2k+1}}{1 + \tau^{2k+1}}\right)^2 + z}{1 + \left(\dfrac{1 - \tau^{2k+1}}{1 + \tau^{2k+1}}\right)^2 z} \tag{6.1}$$

or

$$\begin{aligned}\phi(z; m) &= \prod_{k=0}^{\infty} \frac{(1 - \tau^{2k+1})^2 + z(1 + \tau^{2k+1})^2}{(1 + \tau^{2k+1})^2 + z(1 - \tau^{2k+1})^2} \\ &= \prod_{k=0}^{\infty} \frac{(1 + z) - 2\tau^{2k+1}(1 - z) + \tau^{4k+2}(1 + z)}{(1 + z) + 2\tau^{2k+1}(1 - z) + \tau^{4k+2}(1 + z)} \\ &= \prod_{k=0}^{\infty} \frac{1 - 2\tau^{2k+1}\left(\dfrac{1 - z}{1 + z}\right) + \tau^{4k+2}}{1 + 2\tau^{2k+1}\left(\dfrac{1 - z}{1 + z}\right) + \tau^{4k+2}},\end{aligned}$$

and by (5.9)

$$\phi(z; m) = \theta_4(u)/\theta_3(u) \tag{6.2}$$

where τ is the parameter of both $\theta_4(u)$ and $\theta_3(u)$ and $\cos 2\pi u = (1 - z)/(1 + z)$.

7. A hyperbolic analogue of the Tchebycheff polynomials. It has been observed in **5** that $G[B(z; m)]$ is completely ramified over m and $-m$, having only branch points of order one over these points. Recall the structure of the Riemannian image of the finite z-plane with respect to $w = \cos z$, which is completely ramified over $+1$ and -1 with branch points of order one, and which is locally simple over every other point of the finite w-plane. On the strength of the analogy between the structure of $G[B(z; m)]$ and the Riemann surface for $\cos z$, it is natural to conjecture that in many respects $B(z; m)$ plays much the same role in the theory of functions which are analytic and bounded in the interior of the unit circle that $\cos z$ plays in the theory of entire functions. The similarity of the roles played by $B(z; m)$ and $\cos z$ will be brought out in two extremal problems to be considered. The first is

PROBLEM. *To determine among all $(1, n)$ directly conformal maps, $\pi_n(z)$, of $|z| < 1$ onto itself those for which*

$$\max_{-r \leqq x \leqq r} |\pi_n(x)|$$

is least and the corresponding écart. Here n is a given positive whole number and r is a given positive number less than one.

This problem is a hyperbolic analogue of the problem of determining the Tchebycheff polynomials from their extremal properties.

The first question that must be settled is that of existence. Let $\mu(n)$ denote $\underset{\{\pi_n(z)\}}{\text{g.l.b.}}\,[\max_{-r\leqq x\leqq r}|\pi_n(x)|]$. The class of functions consisting of all constants of modulus one and all directly conformal maps of $|z|<1$ onto itself which are of degree not exceeding n is compact. Hence it may be shown by a familiar argument that there exists a member of this class, $\pi^*_k(z)$, of degree k $(0\leqq k\leqq n)$ for which

$$(7.1)\qquad \max_{-r\leqq x\leqq r}|\pi^*_k(x)|=\mu(n).$$

The degree k cannot be less than n. Otherwise the function $z^{n-k}\pi^*_k(z)$ would belong to the class $\{\pi_n(z)\}$ and would be such that

$$\max_{-r\leqq x\leqq r}|x^{n-k}\pi^*_k(x)|<\mu(n).$$

Hence there exists an extremal function in the class $\{\pi_n(z)\}$ for the present problem. A typical extremal function will henceforth be denoted by $\pi^*_n(z)$.

Qualitative properties of a typical extremal function may be readily determined. Of central importance among such properties for determining $\mu(n)$ and the extremal functions themselves are: 1° the location and the nature of the zeros, 2° the location of the points on $(-r\leqq x\leqq r)$ where $|\pi^*_n(z)|$ attains $\mu(n)$. Recall that a $(1,n)$ directly conformal map of $|z|<1$ onto itself may be represented by

$$(7.2)\qquad \pi_n(z)\equiv e^{i\theta}\prod_{k=1}^{n}\frac{z-\alpha_k}{1-\bar{\alpha}_k z}\qquad(\theta \text{ real};\ |\alpha_k|<1,\ k=1,2,\cdots,n).$$

With the aid of this representation it can be shown that all the zeros of $\pi^*_n(z)$ are real. This will be demonstrated by using the inequality

$$(7.3)\qquad \left|\frac{x-\Re\alpha}{1-(\Re\alpha)x}\right|<\left|\frac{x-\alpha}{1-\bar{\alpha}x}\right|\qquad(|\Re\alpha|<|\alpha|<1)$$

where x is real and satisfies $-1<x<1$. The inequality may be established by direct analytical verification or else by noting that the non-Euclidean circle centered at x and tangent to the Euclidean straight line $\Re z=\Re\alpha$ does not contain α in its closed interior. On referring to the representation (7.2) one can see that by virtue of (7.3) if any zero of $\pi^*_n(z)$ is not real, $\mu(n)$ would be diminished if that zero were replaced by its real part.

Furthermore, every zero of $\pi^*_n(z)$ must lie in $(-r<x<r)$. Suppose that α, a zero of $\pi^*_n(z)$, satisfied $|\alpha|\geqq r$. Then, if α were replaced by $\alpha'=\rho\alpha$ where ρ is real and less than but near one, the value of $\mu(n)$ would be

diminished. Also the zeros of $\pi^*_n(z)$ are all simple. To see this suppose that β is a multiple zero of $\pi^*_n(z)$. Let β_1 and β_2 satisfy: 1° $-r < \beta_1 < \beta < \beta_2 < r$; 2° $\max\limits_{\beta_1 < x < \beta_2} (x - \beta_1)/(1 - \beta_1 x) \cdot (\beta_2 - x)/(1 - \beta_2 x) < \mu(n)$; 3° $(\beta - \beta_1)/(1 - \beta\beta_1) = (\beta_2 - \beta)/(1 - \beta\beta_2)$. If the alleged factor, $((z - \beta)/(1 - \beta z))^2$, of $\pi^*_n(z)$ were replaced by $(z - \beta_1)/(1 - \beta_1 z) \cdot (z - \beta_2)/(1 - \beta_2 z)$, the resulting function would be such that its *écart* on the interval $(-r \leqq x \leqq r)$ would be less than $\mu(n)$. This is certainly true for $\beta_1 \leqq x \leqq \beta_2$ by the choice of β_1 and β_2. On the other hand, for those x which satisfy either $-r \leqq x < \beta_1$ or $\beta_2 < x \leqq r$ the readily verified inequality

$$0 < \frac{x - \beta_1}{1 - \beta_1 x} \cdot \frac{x - \beta_2}{1 - \beta_2 x} < \left(\frac{x - \beta}{1 - \beta x}\right)^2 \tag{7.4}$$

prevails. Hence the *écart* on the interval $(-r \leqq x \leqq r)$ of the function replacing $\pi^*_n(z)$ would be less than $\mu(n)$, and this is contrary to the definition of $\mu(n)$.

If $\pi^*_n(z)$ is subjected to the normalization, $\pi^*_n(1) = 1$, then $\pi^*_n(z)$ is real for z real. Since $\pi^*_n(z)$ is of degree n, all the zeros of $d\pi^*_n/dz$ in $|z| < 1$ are accounted for by the points of relative maxima of $|\pi^*_n|$ between successive zeros of π^*_n, there being precisely one zero of $d\pi^*_n/dz$ between successive zeros of π^*_n. The value of $|\pi^*_n(z)|$ at these positions of relative maxima and minima as well as at $z = -r, r$ is $\mu(n)$. Suppose, for example, that $\pi^*_n(r) < \mu(n)$ (it should be observed that $\pi^*_n(r) > 0$ by virtue of the normalization to which $\pi^*_n(z)$ is subjected) and let the zeros of $\pi^*_n(z)$ be denoted by $\alpha_1, \alpha_2, \cdots, \alpha_n$ where

$$-r < \alpha_1 < \alpha_2 < \cdots < \alpha_n < r.$$

If α_n were replaced by α'_n in $\pi^*_n(z)$, where α'_n is real, less than and sufficiently close to α_n, then the resulting function would be such that the maximum of its modulus on the interval $(-r \leqq x \leqq r)$ would be actually less than $\mu(n)$ and this is impossible. A similar argument would apply if $|\pi^*_n(-r)| < \mu(n)$. There remains for consideration the relative maximum of $|\pi^*_n(x)|$ between α_k and α_{k+1} $(k = 1, 2, \cdots, n-1)$. Let γ_k denote the zero of $d\pi^*_n/dz$ lying between α_k and α_{k+1} $(k = 1, 2, \cdots, n-1)$, and suppose that $|\pi^*_n(\gamma_{k_0})| < \mu(n)$ for some index k_0. Replace α_{k_0} and α_{k_0+1} by α'_{k_0} and α''_{k_0+1} respectively in $\pi^*_n(z)$, where α'_{k_0} and α''_{k_0+1} satisfy the following conditions: 1° $\alpha'_{k_0} < \alpha_{k_0} < \alpha_{k_0+1} < \alpha''_{k_0+1}$; 2° α'_{k_0} is so close to α_{k_0} and α''_{k_0+1} is so close to α_{k_0+1} that the maximum of the modulus of the replacing function for $\alpha'_{k_0} < x < \alpha''_{k_0+1}$ is less than $\mu(n)$; 3° the non-Euclidean distance between α'_{k_0} and α_{k_0} is equal to the non-Euclidean distance between α_{k_0+1} and α''_{k_0+1}. The effect of 3° is that

$$(7.5) \qquad 0 < \frac{x-\alpha'_{k_0}}{1-\alpha'_{k_0}x} \cdot \frac{x-\alpha''_{k_0+1}}{1-\alpha''_{k_0+1}x} < \frac{x-\alpha_{k_0}}{1-\alpha_{k_0}x} \cdot \frac{x-\alpha_{k_0+1}}{1-\alpha_{k_0+1}x}$$

for $-1 < x < \alpha'_{k_0}$ and $\alpha''_{k_0+1} < x < 1$. The function replacing $\pi^*_n(z)$ under this change of zeros would be of modulus less than $\mu(n)$ for $-r \leqq x \leqq r$, which is impossible. To sum up,

THEOREM 7. 1. *There exists a* $(1, n)$ *directly conformal transformation of* $|z| < 1$ *onto itself such that the maximum of its modulus on the interval* $(-r \leqq x \leqq r)$ $(0 < r < 1)$ *is equal to* $\mu(n) = \underset{\{\pi_n(z)\}}{\text{g.l.b.}}[\max_{-r\leqq x\leqq r} |\pi_n(x)|]$ *where* $\pi_n(z)$ *denotes a* $(1, n)$ *directly conformal map of* $|z| < 1$ *onto itself. Further, if* $\pi^*_n(z)$ *denotes an extremal function for this problem so normalized that* $\pi^*_n(1) = 1$, *then* $\pi^*_n(z)$ *has the following properties*:

1° $\pi^*_n(z)$ *is real for* z *real*;

2° *the* n *zeros of* $\pi^*_n(z)$ *are all simple and lie in* $(-r < x < r)$;

3° $\pi^*_n(r) = \mu(n)$, $|\pi^*_n(-r)| = \mu(n)$, *the maximum of* $|\pi^*_n(x)|$ *between successive zeros of* $\pi^*_n(z)$ *is equal to* $\mu(n)$, *this value being attained by* $|\pi^*_n(x)|$ *at* $r, -r$, *and the zeros of* $d\pi^*_n/dz$;

4° *the zeros of* $d\pi^*_n/dz$ *are simple, and there is precisely one zero of* $d\pi^*_n/dz$ *between successive zeros of* $\pi^*_n(z)$, *and there are no others.*

It should be remarked that $|\pi^*_n(z)| = \mu(n)$ only at those points enumerated in 3° since $\pi^*_n(z)$ is of degree n. The structure of the Riemannian image of the extended plane with respect to $w = \pi^*_n(z)$ is completely specified by Theorem 7. 1. Since the derivative of $\pi^*_n(z)$ at $z = 1$ is positive, it follows from the uniqueness theorem of the theory of the conformal mapping of simply-connected Riemann surfaces that $\pi^*_n(z)$ is uniquely defined by its normalization and its extremal property. Hence

THEOREM 7. 2. *The normalized extremal function of Theorem 7. 1 is unique and hence is odd if* n *is odd and even if* n *is even. All the extremal functions are given by* $e^{i\theta}\pi^*_n(z)$ $(\theta$ *real*).

There remains to be considered the determination of $\mu(n)$ and $\pi^*_n(z)$. They may be found by the following argument.[7] Consider the function $\Phi(z)$ defined by

$$(7.7) \qquad \Phi(z) \equiv \pi^*_n[B(z; r)].$$

An analysis of $\Phi(z)$ reveals that $\Phi(z) \equiv B[z; \mu(n)]$.

---

[7] It is readily seen that $\mu(1) = r$ and $\pi^*_1(z) \equiv z$. It will be assumed for the remainder of this section that $n > 1$.

This may be seen by first noting that $B(0;r)=r$ implies $\Phi(0)=\mu(n)$. Second, the structure of $G[\Phi]$ may be readily inferred from the structures of $G[B(z;r)]$ and $G[\pi^*_n]$. From the definition of $B(z;r)$ as $\phi(-z^2;r)$ (5), $G[B(z;r)]$ may be constructed with the aid of a set of simple open disks, S_k, $(k=0,1,2,\cdots;-1,-2,\cdots)$, where each S_k consists of a copy of $|w|<1$ slit along the real axis from -1 to $-r$ and from r to 1. In order to obtain $G[B(z;r)]$ the upper and lower banks of the right slit of S_k are joined to the lower and upper banks respectively of the right slit of S_{k+1}, and the upper and lower banks of the left slit of S_k are joined to the lower and upper banks respectively of the left slit of S_{k-1} for $k=0,\pm 2,\pm 4,\cdots$, the Riemannian domain so obtained being $G[B(z;r)]$. From the descriptive properties of $\pi^*_n(z)$ developed in the present section $G[\pi^*_n]$ may also be constructed. Let σ_λ $(\lambda=1,\cdots,n)$ denote a copy of the open disk, $|w|<1$, slit along the real axis from -1 to $-\mu(n)$ and from $\mu(n)$ to 1. Then $G[\pi^*_n]$ may be formed from the σ_λ as follows. The upper and lower right banks of σ_1 are joined to each other; then the upper and lower left banks of σ_1 are joined to the lower and upper banks respectively of σ_2. If $n=2$, the upper and lower right banks of σ_2 are joined to each other. If $n>2$, then the upper and lower right banks of σ_2 are joined to the lower and upper banks respectively of σ_3, etc. In σ_n one slit remains and the upper and lower banks of this slit are to be joined together. If n is odd, it is the left-hand slit, otherwise the right-hand slit. The validity of this construction is guaranteed by the general properties of $(1,n)$ directly conformal maps of the interior of the unit circle onto itself and the particular ramification properties of $\pi^*_n(z)$ already enumerated.

It follows from this discussion that the Riemannian image of each S_k with respect to $\pi^*_n(z)$ consists of a copy of $G[\pi^*_n]$ with σ_1 slit from $\mu(n)$ to 1 and σ_n slit from $\mu(n)$ to 1 if n is even, or from -1 to $-\mu(n)$ if n is odd. Hence if $\bar{G}^{(k)}[\pi^*_n]$ denotes the Riemannian image of S_k with respect to $\pi^*_n(z)$, $G[\Phi]$ may be constructed in terms of the $\bar{G}^{(k)}[\pi^*_n]$ by connecting the latter along their slits in the following manner. Let it be agreed that the right slit of $\sigma_1{}^{(0)}$ of $\bar{G}^{(0)}[\pi^*_n]$ ($\sigma_\lambda{}^{(k)}$ having the connotation for $\bar{G}^{(k)}[\pi^*_n]$ that σ_λ has for $G[\pi^*_n]$) is to be connected to the right slit of $\sigma_1{}^{(1)}$ and that the as yet unconnected slit of $\sigma_n{}^{(0)}$ be connected to the corresponding slit of $\sigma_n{}^{(-1)}$, upper and lower banks being connected with lower and upper banks respectively. The next step is to connect the available slit of $\sigma_n{}^{(1)}$ to the corresponding slit of $\sigma_n{}^{(2)}$ and the right slit of $\sigma_1{}^{(-1)}$ to the right slit of $\sigma_1{}^{(-2)}$, again with the same type of connection between upper and lower banks. The process is to be continued in this way step-by-step. In this manner it is seen that

$G[\Phi]$ and $G[B(z;\mu(n))]$ are identical. Thus, since $\Phi(0) = B(0;\mu(n))$, $\Phi(z) \equiv B(e^{i\theta}z;\mu(n))$ (θ real). Now $\theta \equiv 0 \pmod{\pi}$; this may be established by noticing that the zeros of $\Phi(z)$ are to be found from those of $B(z;\mu(n))$ by multiplying those of the latter by $e^{-i\theta}$; all the zeros of $\Phi(z)$ and $B(z;\mu(n))$ being real, the assertion follows. Since $B(z;\mu(n))$ is even, $\Phi(z) \equiv B(z;\mu(n))$.

The following argument determines $\mu(n)$ and the zeros of $\pi^*_n(z)$. Let T_r denote the linear fractional transformation of $|z| < 1$ onto itself associated with $B(z;r)$ as T is associated with $B(z;m)$ in **5** and let λ_r denote the associated multiplier (cf. (5.2)); similarly let $T_{\mu(n)}$ and $\lambda_{\mu(n)}$ have the corresponding connotations for $B(z;\mu(n))$. From (7.7) it follows that the cyclic group generated by T_r is a subgroup of the cyclic group generated by $T_{\mu(n)}$. Hence $T_r = T^k{}_{\mu(n)}$ for some positive integer k. Actually $k = n$. This may be seen by noting that $B(z;\mu(n))$ has precisely k simple zeros in the interval $(0 < x < T^k{}_{\mu(m)}0)$ or, what is the same, in the interval $(0 < x < T_r 0)$. Now $B(x;r)$ is monotone decreasing for $(0 \leqq x \leqq T_r 0)$ and $B(0;r) = r$ and $B(T_r 0;r) = -r$. Since the n zeros of $\pi^*_n(z)$ are located in the open interval $(-r, r)$, $\pi^*_n[B(x;r)] \equiv B(x;\mu(n))$ has n simple zeros in $(0 < x < T_r 0)$. Hence $k = n$. A consequence of this result is that $\lambda_{\mu(n)} = \lambda_r{}^{1/n}$. On reference to (5.6), the value $\mu(n)$ may be obtained at once. It is

$$\mu(n) = \prod_{k=0}^{\infty} \left[\frac{1 - \lambda_r{}^{(2k+1)/2n}}{1 + \lambda_r{}^{(2k+1)/2n}}\right]^2. \tag{7.8}$$

Also the Blaschke product for $B(z;\mu(n))$ may be written down at once with the aid of (5.4) and (5.5).

The extremal function, $\pi^*_n(z)$, will be "specified" once its zeros are given. These may be found from the identity already employed:

$$B(z;\mu(n)) \equiv \pi^*_n[B(z;r)]. \tag{7.9}$$

On the interval $(0 < x < T_r 0)$, $B(z;\mu(n))$ has n simple zeros. In fact, they are $T_r{}^{(2k-1)/2n}0$ $(k = 1, 2, \cdots, n)$. Since $B(x;r)$ is (1,1) on the interval $(0 \leqq x \leqq T_r 0)$, the points $B[T^{(2k-1)/2n}0;r]$ $(k = 1, 2, \cdots, n)$ are distinct, lie in the interval $(-r < x < r)$ and hence by (7.9) are the zeros of $\pi^*_n(z)$ and there are no others. The argument is complete.

Remark I. This discussion may be paraphrased for the theory of ordinary Tchebycheff polynomials with the starting point of the discussion being the definition of the Tchebycheff polynomial as one which enjoys the extremal property usually associated with it rather than the *ad hoc* definition

$$T_n(x) \equiv \cos(n \arccos x)/2^n.$$

Remark II. It should be noted that the class of competing functions $\{\pi^*_n(z)\}$ is not convex. Hence the useful property of convexity available in the theory of ordinary Tchebycheff polynomials may not be employed here. The full force of the proof is to be found in the representation (7.2) of a competing function, in the associated discussion of the location of the zeros of an alleged extremal function, and in the ramification properties of the mapping it defines.

8. An extremal problem for functions analytic and of modulus less than one in the interior of the unit circle, which vanish at the origin and have an assigned derivative there. The second extremal problem alluded to in 7 is:

PROBLEM. *Let* E_α *denote the class of functions,* $f(z)$, *which are analytic and of modulus less than one for* $|z| < 1$ *and satisfy the conditions:* 1° $f(0) = 0$, 2° $f'(0) = \alpha$, *where* α *is an assigned real positive number less than one. Further let* $\mu(\alpha) = \underset{f \in E_\alpha}{\text{g.l.b.}} [\underset{-1<x<1}{\text{l.u.b.}} |f(x)|]$. *To determine* $\mu(\alpha)$ *and the corresponding extremal function.*[8]

This problem may be solved in terms of the function $B(z; m)$ by a proper choice of m. Since the solution of the present problem calls for many details, the major steps are outlined.

I. *There exists an extremal function for this problem.* This may be inferred from the compactness of the class E_α. Let $\phi(z)$ denote a typical extremal function. Actually it will turn out that there is a unique extremal function.

II. $|\phi(z)|$ *attains* $\mu(\alpha)$ *for some* z *in the open interval of the real axis* $(-1 < x < 1)$.

III. *In fact,* $|\phi(z)|$ *attains* $\mu(\alpha)$ *an infinite number of times on the interval* $(-1 < x < 1)$.

IV. *The extremal function is unique.* From this stage of the solution on, $\phi(z)$ is unambiguous in connotation.

V. $\phi(z)$ *is real for* z *real and is odd.*

VI. *All the zeros of* $\phi(z)$ *are real.*

VII. At this juncture it is convenient to introduce an auxiliary extremal problem. Let $\mu(\alpha, r) = \underset{f \in E_\alpha}{\text{g.l.b.}} [\underset{-r \leqq x \leqq r}{\max} |f(x)|]$, and let it be required to determine properties of the associated extremal function. (The problem of determining $\mu(\alpha, r)$ and the corresponding extremal function explicitly is difficult but its solution is not essential for the solution of the major problem at hand.) The following properties of the extremal function for this subsidiary problem are listed without proof.[9] (a) *From the compactness of* E_α, *it follows that there exists an extremal function.* (b) *The extremal function is unique. It defines a* $(1, k(r))$ ($k(r)$ *a positive integer*) *directly conformal map of* $|z| < 1$ *onto itself. Let it be denoted by* $\psi(z; r)$. (c) $\psi(z; r)$ *is real for* z *real and is odd.* (d) *The zeros of* $\psi(z; r)$ *are real and simple.* (e) *At most one zero of* $\psi(z; r)$ *lies to the right of* $x = r$ *and at most one zero of* $\psi(z; r)$ *to the left of* $x = -r$. (f) $\mu(\alpha, r)$ *is attained on the real axis by* $|\psi(z; r)|$ *between successive zeros of* $\psi(z; r)$. (g) *The antecedent of the closed interval* $[-\mu(\alpha, r), \mu(\alpha, r)]$ *with respect to* $w = \psi(z; r)$ *is a subset of* $(-1 < x < 1)$. (h) $\mu(\alpha, r)$ *is monotone non-decreasing for* $0 < r < 1$ *and* $\lim_{r \to 1} \mu(\alpha, r) = \mu(\alpha)$. *Hence by* IV $\lim_{r \to 1} \psi(z; r) = \phi(z)$ *in the sense of continuous convergence for* $|z| < 1$.

---

[8] Cf. Bernstein [1] where a related problem for entire functions is treated.

VIII. From VII (g) and (h) it may be proved that *the antecedent of the closed interval* $[-\mu(\alpha), \mu(\alpha)]$ *with respect to* $w = \phi(z)$ *is* $(-1 < x < 1)$.

IX. *At each point of* $(-1 < x < 1)$ *where* $|\phi| = \mu(\alpha)$, ϕ' *has a zero of order one.*

X. From IX it is concluded that $\phi(z) = B[T^{-\frac{1}{2}}z; \mu(\alpha)]$ *where* T *has the meaning previously associated with it.*

XI. $\mu(\alpha)$ *is determined in terms of* α *from the relation* $\phi'(0) = \alpha$.

The proof of I will be omitted inasmuch as the argument is standard. To establish II the auxiliary function, $\eta(z; \beta)$, is introduced, being defined by

$$\eta(z; \beta) \equiv z e^{\beta(z^2+1)/(z^2-1)} \quad (\beta \text{ real and positive}). \tag{8.1}$$

With this restriction on β, $\eta(z; \beta)$ is analytic and of modulus less than one for $|z| < 1$ and in addition has the property that $\lim_{x \to 1} \eta(x; \beta) = \lim_{x \to -1} \eta(x; \beta) = 0$, where x is assumed real and of modulus less than one. If β is taken equal to $-\log \alpha$, then $\eta(z; \beta) \in E_\alpha$. Let it be assumed for the immediate argument that this is the case. The statement II may be established as follows. Suppose that $\mu(\alpha)$ is not attained by $|\phi(z)|$ on the open interval $(-1 < x < 1)$ and let r be chosen positive and less than, but so close to, one that for x real and satisfying $r \leqq |x| < 1$

$$|\eta(x; \beta)| \leqq \mu(\alpha)/2. \tag{8.2}$$

---

[9] The proofs of these properties may be supplied by applying the methods of § 7 of [5]. Since the proofs of the present results are so closely related to those given there, they will be omitted.

Clearly $\max\limits_{-r \leqq x \leqq r} | \phi(x) | < \mu(\alpha)$. Now consider

$$\Phi(z;\lambda) \equiv (1-\lambda)\phi(z) + \lambda\eta(z;\beta) \tag{8.3}$$

for λ real and satisfying $0 \leqq \lambda \leqq 1$. For such λ the function (8.3) belongs to E_α. For λ positive and sufficiently small, $\Phi(z;\lambda)$ satisfies (i) $\max\limits_{-r \leqq x \leqq r} | \Phi(x;\lambda) | < \mu(\alpha)$; (ii) $\operatorname*{l.u.b.}\limits_{r<|x|<1} | \Phi(x;\lambda) | < \mu(\alpha)$ (x real). Hence $\operatorname*{l.u.b.}\limits_{-1<x<1} | \Phi(x;\lambda) | < \mu(\alpha)$ and this contradicts the assumption that $\phi(z)$ is extremal. II follows.

The assertions III and IV may be established by a combined argument. Let $\{x_k\}$ denote the set of points on $(-1 < x < 1)$ for which $| \phi(x_k) | = \mu(\alpha)$ and consider the Pick-Nevanlinna interpolation problem[10] of determining the functions, $g(z)$, which are analytic and of modulus less than one for $| z | < 1$ and satisfy in addition

$$g(x_k) = \phi(x_k), \qquad g(0) = 0 \tag{8.4}$$

for all indices k considered. The following inferences may be drawn from the Pick-Nevanlinna interpolation theory. The case where the set $\{x_k\}$ is finite will be treated first. If the function $g(z)$ is uniquely determined by the interpolation conditions (8.4), then it is rational and defines a $(1, K)$ directly conformal map of $| z | < 1$ onto itself for some positive integer K. This is impossible because then $\phi(z)$ would have the property

$$\lim_{|z| \to 1} | \phi(z) | = 1 \tag{8.5}$$

which is contrary to its definition as an extremal function for the major problem under consideration.

If $g(z)$ is not uniquely determined by (8.4), then the possible values of $g'(0)$ for the class of $g(z)$ considered fill a closed circle containing α. Let ρ (< 1) denote the largest positive number in this circle. Clearly, $\alpha \leqq \rho$. There is a unique member of the class of $g(z)$, say $g_0(z)$, which satisfies the further condition that $g'_0(0) = \rho$; $g_0(z)$ defines a $(1, K)$ directly conformal map of $| z | < 1$ onto itself. The assumption, $\alpha = \rho$, is untenable by the argument used above. If $\alpha < \rho$, take β in $\eta(z;\beta)$ equal to $-\log \alpha/\rho$. Then the function

$$\phi^*(z) \equiv g_0(z)\eta(z;\beta)/z \tag{8.6}$$

belongs to E_α and is such that for the x_k considered

$$| \phi^*(x_k) | < \mu(\alpha). \tag{8.7}$$

---

[10] Cf. Walsh [15], R. Nevanlinna [13].

Therefore for λ sufficiently small and positive, $(1-\lambda)\phi(z)+\lambda\phi^*(z)$ belongs to E_a and satisfies

$$(8.8) \qquad \underset{-1<x<1}{\text{l.u.b.}} |(1-\lambda)\phi(x)+\lambda\phi^*(x)| < \mu(\alpha)$$

which contradicts the extremal character of $\phi(z)$. *Therefore an extremal,* $\phi(z)$, *satisfies* III.

The situation where the set $\{x_k\}$ is infinite remains to be considered. Again, if $g(z)$ is not uniquely determined by (8.4), the set of possible values of $g'(0)$ fills a closed circle, containing α. Let ρ^* (<1) denote the largest positive number in this circle. The possibility that $\alpha<\rho^*$, is excluded by the argument applied to the case where the set $\{x_k\}$ is finite. Hence $\alpha=\rho^*$, and this implies, by the Pick-Nevanlinna theory, that $g(z)$ is uniquely determined by (8.4) and the condition $g'(0)=\alpha$. There remains only the case where $g(z)$ is uniquely determined. At all events the only function analytic and of modulus less than one for $|z|<1$ which satisfies (8.4) and $g'(0)=\alpha$ is $\phi(z)$ itself.

Now suppose that $\phi(z)$ is not unique and let $\phi_1(z)$ and $\phi_2(z)$ denote two distinct extremal functions of class E_a. Clearly, $[\phi_1(z)+\phi_2(z)]/2$ also belongs to E_a, is extremal, and is distinct from ϕ_1 and ϕ_2. Let $\{X_k\}$ denote the infinite set of points on $(-1<x<1)$ for which $|\phi_1+\phi_2|/2=\mu(\alpha)$. At such points $\phi_1=(\phi_1+\phi_2)/2=\phi_2$. Hence by the result of the preceding paragraph

$$(8.9) \qquad \phi_1(z)\equiv\frac{\phi_1(z)+\phi_2(z)}{2}\equiv\phi_2(z)$$

which is impossible; uniqueness is established.

Assertion V is readily proved by noting that along with $\phi(z)$ both $\overline{\phi(\bar z)}$ and $-\phi(-z)$ are extremal.

Assertion VI may be concluded in the same manner as VII (d). The proof is readily supplied.

Assertion VIII may be established as follows. Suppose that there exists a point z_0 of $|z|<1$, not on the interval $(-1<x<1)$, with $\phi(z_0)$ in the closed interval $[-\mu(\alpha),\mu(\alpha)]$. Then there would exist in $|z-z_0| < \min[|Iz_0|, 1-|z_0|]$ a point ζ such that $\phi(\zeta)$ would lie in the closed interval $[-\mu(\alpha,r),\mu(\alpha,r)]$ for all r less than and sufficiently near one. But then by Hurwitz's theorem, for a given neighborhood of ζ, say $N(\zeta)$, and for r sufficiently near one, $\psi(z;r)$ would attain the value $\phi(\zeta)$ for some z in $N(\zeta)$ and this manifestly contradicts VII (g).

As for IX, the points of $(-1<x<1)$ where $|\phi(x)|=\mu(\alpha)$ are points where $\phi(x)$ has either a relative maximum or minimum, $\phi(x)$ being considered as a real function of the real variable x. At such points ϕ' has a zero of order one as a consequence of VIII.

Step X is decisive in completing the proof. Let $b(z)$ denote $B(T^{-\frac{1}{2}}z; \mu(\alpha))$ and consider the equation

$$(8.10) \qquad b[\psi(z)] = \phi(z), \qquad (\psi(0) = 0).$$

Since $b(0) = 0$, it follows that (8.10) defines $\psi(z)$ as a single-valued analytic function of modulus less than one for $|z| < 1$, since $G[\phi]$ is completely ramified with branch points of order one over $\mu(\alpha)$ and $-\mu(\alpha)$. Upon differentiation, (8.10) yields $b'(0)\psi'(0) = \phi'(0)$. By Schwarz's lemma $|\psi'(0)| \leqq 1$. Hence $0 < \alpha = \phi'(0) \leqq b'(0)$. It is impossible that $\alpha < b'(0)$, since then $(\alpha/b'(0))b(z)$ would belong to E_α and yet for real z, $|(\alpha/b'(0))b(z)| \leqq (\alpha/b'(0))\mu(\alpha) < \mu(\alpha)$. This would imply that $\phi(z)$ is not extremal. Therefore since $\alpha = b'(0)$, it follows that $\phi(z) \equiv b(z)$.

There remains to be considered the determination of $\mu(\alpha)$ and the multiplier of T in terms of α (XI). These questions may be treated by much the same methods as those used in **5**. As in that section let T be given by

$$(8.11) \qquad T: \frac{Z-1}{Z+1} = \lambda \frac{z-1}{z+1} \qquad (0 < \lambda < 1).$$

Then $b(z)$ may be represented by

$$(8.12) \qquad b(z) \equiv z \prod_{k=1}^{\infty} \frac{(T^k 0)^2 - z^2}{1 - (T^k 0)^2 z^2}.$$

Hence $\alpha = \prod_{k=1}^{\infty} [T^k 0]^2$. But $(T^k 0)^2 = \left(\frac{1-\lambda^k}{1+\lambda^k}\right)^2$. It follows that

$$(8.13) \qquad \alpha = \prod_{k=1}^{\infty} \left(\frac{1-\lambda^k}{1+\lambda^k}\right)^2.$$

It is readily seen as in **5** that there exists a unique λ $(0 < \lambda < 1)$ satisfying (8.13). With λ so determined, $\phi(z)$ is given by (8.12). The extremal value, $\mu(\alpha)$, is $b(T^{\frac{1}{2}}0)$. The solution of the problem is complete.

Remark. Related extremal problems may be treated in the same spirit. For example, one may consider the problem of determining

$$\underset{f \in E_\alpha}{\text{g.l.b.}} \, [\underset{0 < x < 1}{\text{l.u.b.}} \, |f(x)|]$$

and the associated extremal function. The problem treated in this section and the one just alluded to are of interest inasmuch as their solutions yield precise information on the magnitude of a function of class E_α either on a diameter of the unit circle or a ray issuing from the origin. Thus an easy corollary of the present discussion is

COROLLARY. *If* $f(z)$ *is analytic and of modulus less than one for* $|z| < 1$, *and if further* $f(0) = 0$, $|f'(0)| = \alpha$ $(0 < \alpha < 1)$, *then along each diameter of* $|z| < 1$ *there exists a point at which* $|f| \geqq \mu(\alpha)$. *If along any diameter* $|f| \leqq \mu(\alpha)$, *then* $f(z) \equiv e^{i\theta_1}\phi(e^{i\theta_2}z)$, *where* θ_1 *and* θ_2 *are real constants.*

ILLINOIS INSTITUTE OF TECHNOLOGY.

BIBLIOGRAPHY.

1. S. Bernstein, *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Paris, 1926.
2. A. Beurling, *Études sur un problème de majoration*, Thesis, Upsala, 1933.
3. G. C. Evans, *The logarithmic potential*, New York, 1927.
4. W. Fenchel, "Ueber analytische Funktionen, die in Teilgebieten des Einheitskreises beschränkt sind," *Nachrichten Göttingen* (1930), pp. 124-132.
5. M. H. Heins, "On a problem of Walsh concerning the Hadamard three circles theorem," *Transactions of the American Mathematical Society*, vol. 55 (1944), pp. 349-372.
6. G. Herglotz, "Ueber Potenzreihen mit positivem, reellem Teil im Einheitskreis," *Leipziger Berichte*, vol. 63 (1911).
7. G. Julia, *Principes géométriques d'analyse*, Première partie, Paris, 1930.
8. E. Landau, "Ueber den Millouxschen Satz," *Nachrichten Göttingen* (1930), pp. 1-9.
9. H. Milloux, "Le théorème de M. Picard, suites de fonctions holomorphes, fonctions méromorphes et entières," *Journal des Mathématiques pures et appliquées*, vol. 9, s. 3 (1924).
10. H. Milloux, "Sur certaines fonctions holomorphes et bornées dans un cercle," *Mathematica*, vol. 4, pp. 182-185.
11. R. Nevanlinna, *Eindeutige analytische Funktionen*, Berlin, 1936.
12. R. Nevanlinna, "Ueber eine Minimumaufgabe in der Theorie der konformen Abbildung," *Göttingen Nachrichten* (37), 1933.
13. R. Nevanlinna, "Ueber beschränkte analytische Funktionen," *Ann. Acad. Sci. Fenn.*, vol. 32, no. 7.
14. E. Schmidt, "Ueber den Millouxschen Satz," *Sitzber. Pr. Akad. Wiss. Phys. Math. Kl.*, 1932, pp. 394-401.
15. J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, New York, 1935.
16. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Fourth Edition, Cambridge, 1927.

THE DENSITIES OF IDEAL CLASSES AND THE EXISTENCE OF UNITIES IN ALGEBRAIC NUMBER FIELDS.*

By AUREL WINTNER.

According to a terminology customary in the analytic theory of numbers, the "depth" of an asymptotic law concerning integers or ideals is measured in terms of the (*specific*) *arithmetical* information needed in the proof. Correspondingly, the varying degree of complication in the (*general*) *analytical* theorems of the Tauberian machinery involved is not taken into account by this principle of classification. In this technical sense, the prime number theorem is deeper than Mertens' asymptotic formula, not because a Tauberian theorem relevant for the former requires a proof which happens to be longer than the proof of a Tauberian theorem sufficient for the latter, but merely because the former does, and the latter does not, depend on a certain arithmetical information; namely, on one involving the location of the zeros of $\zeta(s)$.

For the "elementary" laws of algebraic number fields, it is natural to consider a corresponding principle of classification, by isolating the theorems which can be proved without involving the arithmetical fact represented by the existence of a complete system of unities or of an associated lattice. In particular, the Dirichlet-Dedekind theorem, concerning the ergodic distribution of the integral ideals over the various ideal classes of the field, *appears* to be less "deep" (that is to say less arithmetical) than an asymptotic estimate involving such results as Minkowski's theorems on discriminants. However, it is another matter to *prove* that such is the case. In fact, the classical proof of the law of Dirichlet-Dedekind depends on the arithmetical theorem concerning the existence of a lattice.

Thus it might be of interest for algebraists that they are overestimating precisely the "algebraical depth" of the Dirichlet-Dedekind theorem, since *the theorem on the existence of unities* (*lattices*) *is superfluous in the proof for the existence and the equality of the densities of the various ideal classes in an arbitrary* (*finite*) *algebraic number field.*

The resulting proof is not recommended for class-room purposes. It certainly is not simpler than the classical proof. However, it would be about as meaningless to state that the latter is simpler than the former. In fact, one

---

* Received June 28, 1944.

could say that the classical proof is "simpler" only at the cost of introducing the deep, that is to say arithmetical, circumstance that there exists in the field a complete system of unities, transforming the problem into what Minkowski calls a macroscopic volume-evaluation. But it turns out that, historically, this rôle of the theory of the unities was that of a *deus ex machina* (even though, according to Furtwängler, the theory of algebraic numbers can *to-day* be developed so as to be based on a lattice).

It seems to be of interest that the legitimacy of this methodical point of view is fully supported by the historical development. In fact, when Dirichlet introduced unities into his (quadratic) fields, his sole purpose was the proof of the class-number formulae. But the latter are identical with what results if the Dirichlet-Dedekind relations of an arbitrary field are particularized to the case of a quadratic field. In Kummer's adaptation of Dirichlet's theory to cyclotomic fields, the rôle of the unities appeared in a form less pure, since it was interwoven with the algebraic affairs of Fermat's problem. However, in Dedekind's extension of Dirichlet's work to arbitrary fields, the introduction of the unities served again no useful purpose distinct from the one for which Dirichlet originally devised it in his particular case. For the modern algebraist who cares for generalized theories of unities for the unities' sake, this may not be easy to believe, but he can convince himself by turning over Dedekind's supplements [1] or Minkowski's historical sketch [3].

The ironical twist of historical development, which first makes mathematics devise a theory for a definite purpose and subsequently leads to other concrete ends (this time to Minkowski's geometry of numbers) but ultimately proves that the introduction of the theory was superfluous exactly for the original purpose, is not of course an uncommon occurrence. That it presents itself in case of the Dirichlet-Dedekind theorem, can be proved as follows:

With reference to any ideal class C of a fixed algebraic number field $K = K(\alpha)$, let $f(n; C)$ denote the number of the integral ideals which are contained in C and have a norm not exceeding n, where $n = 1, 2, \cdots$. Then the assertion to be proved is that the ratio $f(n; C)/n$ tends to a finite, non-vanishing limit, say λ; and that, in addition, this limit $\lambda = \lambda_K$ is independent of the choice of C in K.

Let $F(n; C)$ denote the number ($\geqq 0$) of the representations of n as norms of (integral) ideals contained in C. Then it is obvious that

$$f(n; C) = F(1; C) + F(2; C) + \cdots + F(n; C) \tag{1}$$

and that, if $\zeta(s; C)$ denotes the zeta-function of C, the series

$$\zeta(s;C)=\sum_{n=1}^{\infty}F(n;C)/n^s \tag{2}$$

converges in the half-plane $\sigma>1$, where $s=\sigma+it$. It follows that the existence of a positive value $\lambda=\lambda_K$ satisfying

$$(\sigma-1)\zeta(\sigma;C)\rightarrow\lambda \quad \text{as} \quad \sigma\rightarrow 1+0$$

is a necessary condition for the truth of the assertion. In fact, Dirichlet's Abelian lemma states that, if $a_1, a_2, \cdots$ is any sequence of values possessing a finite mean-value

$$M(a)=\lim_{n\rightarrow\infty}(a_1+a_2+\cdots+a_n)/n, \tag{3}$$

then the Dirichlet series

$$\sum_{n=1}^{\infty}a_n/n^s \tag{4}$$

must converge for $\sigma>1$ to a function satisfying

$$(\sigma-1)\sum_{n=1}^{\infty}a_n/n^{\sigma}\rightarrow M(a) \quad \text{as} \quad \sigma\rightarrow 1+0. \tag{5}$$

On the other hand, the expression on the left of the limit relation (5) can tend to a limit and, what is more, the function represented by (4) in the half-plane $\sigma>1$ can have a simple pole at $s=1$, even though the limit (3) fails to exist. This is shown by a well-known example of Dedekind, rediscovered by Hardy, and subsequently used by Ramanujan for a mistaken proof of the prime number theorem in a manner the illegitimacy of which was verified by Dedekind precisely on this example; cf. [5]. It is true that the coefficients $F(n;C)$ of (2) are non-negative, but this is of no avail, since the restriction $a_n\geqq 0$ is satisfied by Dedekind's counter-example also.

However, as a corollary of Hecke's functional equation, the function $(s-1)\zeta(s;C)$ is regular in the whole plane, and the residue of $\zeta(s;C)$ at the point $s=1$ has a positive value, say $\lambda=\lambda_K$, which is independent of the choice of C in K. This information is available for the purpose at hand. In fact, while Hecke's original proof did make use of the existence of a lattice, he subsequently observed that his functional equation can be proved without any reference to the theory of unities, and such proofs are now available in the literature. For older references cf. [4].

Actually, neither the functional equation nor the regularity of the difference

$$\zeta(s;C)-\lambda/(s-1) \tag{6}$$

in the whole plane will be needed in the sequel. All that will be used is that the function (6), where $\sigma > 1$, goes over into a continuous boundary function on the line $\sigma = 1$.

In fact, this in itself makes applicable Ikehara's theorem, which is a particular case of a purely analytical result concerning Laplace transforms and does not presuppose *any* arithmetical information; cf., e. g., [2]. Ikehara's theorem states that, if $a_1, a_2, \cdots$ is a sequence of real, non-negative values, and if the corresponding Dirichlet series (4) converges in the half-plane $\sigma > 1$ to a function corresponding to which there exists a constant λ such that the difference

$$\sum_{n=1}^{\infty} a_n/n^s - \lambda/(s-1), \tag{6 bis}$$

where $\sigma > 1$, goes over into a continuous boundary function on the line $\sigma = 1$, then

$$a_1 + a_2 + \cdots + a_n \sim \lambda n$$

as $n \to \infty$. Hence, if (4) is identified with (2), it follows from (1) that $f(n; C) \sim \lambda n$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. G. L. Dirichlet, *Vorlesungen über Zahlentheorie,* herausgegeben und mit Zusätzen versehen von R. Dedekind, 2nd ed., Braunschweig, 1871.

[2] G. Doetsch, *Theorie und Anwendung der Laplace-Transformation,* Dover Publications, New York, 1943, pp. 216-222.

[3] H. Minkowski, *Gesammelte Abhandlungen,* vol. 2, pp. 453-457.

[4] "Report of the committee [L. E. Dickson, H. H. Mitchell, H. S. Vandiver, G. E. Wahlin] on algebraic numbers," *Bulletin of the National Research Council,* vol. 5, part 3, no. 28 (1923), pp. 21-25.

[5] A. Wintner, "On an elementary analogue of the Riemann-Mangoldt formula," *Bulletin of the American Mathematical Society,* vol. 48 (1942), pp. 759-762.

SOME RELATIONS BETWEEN THE BEHAVIOR OF A FUNCTION AND THE ABSOLUTE SUMMABILITY OF ITS FOURIER SERIES.*

By Kien-Kwong Chen.

1. We suppose that $f(t)$ is a periodic function, with period 2π, integrable in the Lebesgue sense. We write

$$\phi(t) = \tfrac{1}{2}\{f(x+t) + f(x-t)\},$$

and suppose the Fourier series of $\phi(t)$ to be $\Sigma A_n \cos nt$. Bosanquet[1] has shown that the convergence of $\Sigma \mid A_n \mid$ does not imply that the mean-function

$$[\phi(t)]_1 = (1/t) \int_0^t \phi(u)\,du$$

is of bounded variation in $(0, \pi)$. It has, however,[2] been proved that the later property on $[\phi(t)]_1$ is a consequence of

$$\Sigma \mid A_n \mid \log n < \infty.$$

In the following section we prove that the convergence of $\Sigma \mid a_n \mid$ implies that

$$\int_0^\delta \mid (d/dt)\{(\log 1/t)^{-2}[\phi(t)]_1\} \mid dt < \infty, \quad (0 < \delta < 1).$$

This is the case $\rho(t) = \mid \log t \mid^{-2}$, $A = \delta$ of Theorem 1. The chief object of the present paper is to establish Theorem 2 which includes both the theorems stated above.

2. Theorem 1. *If the Fourier series of* $f(t)$ *converges absolutely at* $t = x$ *then the function*

$$\Phi(t) = (\rho(t)/t) \int_0^t \{f(x+t) + f(x-t)\}dt$$

is of bounded variation in $(0, A)$, *where* $\rho(t)$ *denotes any function which is absolutely continuous in* $(0, A)$ *such that*

$$\int_0^A \mid t^{-1}\rho(t) \mid dt < \infty.$$

---

* Received May 1, 1944.

[1] L. S. Bosanquet (1).

[2] K. K. Chen (1).

As stated above, this proposition is involved in Theorem 2, but there is a direct proof much simpler than the proof of Theorem 2.

In fact, if $0 < t < A$ then

$$\Phi'(t) \equiv \rho'(t) \Sigma A_n \frac{\sin nt}{nt} + \rho(t) \Sigma A_n \frac{d}{dt} \frac{\sin nt}{nt}. \tag{1}$$

Since

$$\int_0^A \left| \rho'(t) \frac{\sin nt}{nt} \right| dt \leqq \int_0^A | \rho'(t) | \, dt$$

and

$$\int_0^A \left| \rho(t) \frac{d}{dt} \frac{\sin nt}{nt} \right| dt \leqq 2 \int_0^A | \rho(t) t^{-1} | \, dt,$$

it follows from equation (1) that

$$\int_0^A | \Phi'(t) | \, dt \leqq \left(\int_0^A | \rho'(t) | \, dt + 2 \int_0^A | \rho(t) t^{-1} | \, dt \right) \Sigma | A_n |.$$

This proves the theorem.

3. Write now σ_n^α for the n-th Cesàro mean of order α of the series ΣA_n. The series ΣA_n is said to be super-absolutely summable (C, α) to the sum S with respect to the sequence $l = \{l_n\}$ $(0 < l_n < l_{n+1})$ or summable $l \, | \, C, \alpha \, |$ to S, if $\sigma_n^\alpha \rightarrow S$ and the series

$$\Sigma l_n \, | \, \sigma_n^\alpha - \sigma_{n-1}^\alpha \, |$$

is convergent. We denote this mode of summability by the equation

$$\Sigma A_n = S(l \, | \, C, \alpha \, |). \tag{2}$$

We also write, for an integrable function $\chi(u)$,

$$[\chi(t)]_\alpha = \alpha t^{-\alpha} \int_0^t (t-u)^{\alpha-1} \chi(u) \, du \qquad (\alpha > 0)$$

as the mean function of order α of $\chi(u)$ on $(0, t)$.

The theorem on the summability $\{\log n\} \, | \, C, 0 \, |$ stated in **1** has been extended as follows:[3] if (2) holds true for a sequence $l = \{(\log n)^p\}$ $(p \geqq 1)$ and an order $\alpha > -1$, then the function $| \log t |^{p-1} \cdot t^\lambda [t^{-\lambda} \phi(t)]_{1+\alpha}$ is of bounded variation in $(0, \pi)$, provided that $\lambda < 1$, $\lambda \leqq 1 - (\alpha)$, where (α) denotes $\alpha - [\alpha]$, the fractional part of α. The proof depends upon a number of lemmas, some of which are recalled in the next section as they are relevant to the investigations of the present paper.

---

[3] K. K. Chen (1).

4. LEMMA 1. *Let* $b \geqq \alpha > -1$, $\lambda < 1$, $\lambda \leqq 1 - (\alpha)$, $0 < t \leqq \pi$, *and write*

$$(\beta)_n = \frac{(\beta+1)(\beta+2)\cdots(\beta+n)}{n!}, \qquad C(x) = \int_0^1 u^{-\lambda}(1-u)^b \cos(xu)\,du;$$

then the series

$$\sum_{\mu=\nu}^{\infty} (\alpha)_\mu \sum_{n=\mu}^{\infty} (-2-\alpha)_{n-\mu} C(nt)$$

converges to a function $F_\nu(t)$ *such that*

$$F_\nu(t) = O(\nu t)^{-p} \quad (p\text{: a positive constant}), \qquad tF'_\nu(t) = O(1).$$

Moreover, if $\nu t \leqq 1$, *then*

$$F_\nu(t) = B(1-\lambda, 1+\alpha) + O(\nu t)^{\alpha+1}, \qquad F'_\nu(t) = O(\nu).$$

LEMMA 2. *Under the conditions of Lemma* 1, *we have*

$$\int_0^1 u^{-\lambda}(1-u)^b \phi(ut)\,du = \sum_{\nu=0}^{\infty} (\sigma_\nu^\alpha - \sigma_{\nu-1}^\alpha) F_\nu(t) \qquad (\sigma_{-1}^\alpha = 0),$$

provided that $\Sigma \mid \sigma_n^\alpha - \sigma_{n-1}^\alpha \mid$ *converges.*

5. The method we have chosen for the purpose of establishing the theorem stated in **3** may be generalized. Indeed, we can prove the following theorem.

THEOREM 2. *Let* $\rho(t)$ *be a monotone function of* t *in* $(0, \pi)$, *satisfying the conditions*

$$(3) \qquad 0 < \rho(t) = O(|\log t|^A), \qquad \rho(t) = o(L(t)),$$

as $t \to 0$, *where* A *is a constant, and* $L(t) = \int_t^\pi \rho(u) u^{-1} du$. *If the Fourier series of* $f(t)$, *at the point* $t = x$, *is summable* $l \mid C, \alpha \mid$ *for an order* $\alpha > -1$ *to the sum* 0, *where* $l_n = L(1/n)$, $(n = 1, 2, 3, \cdots)$, *then the function*

$$\Psi(t) \equiv \rho(t) t^\lambda [t^{-\lambda}\phi(t)]_{1+\alpha} \qquad (\lambda < 1,\ \lambda \leqq 1 - (\alpha))$$

is of bounded variation in $(0, \pi)$.

Before proving this theorem it is convenient to give first the proof of the following lemma.

6. LEMMA 3. *If the derivative* $\rho'(t)$ *of* $\rho(t)$ *exists and is continuous for* $0 < t \leqq \pi$, *then the conditions* (3) *involve*

$$(4) \qquad (1/t) \int_0^t \rho(u)\,du = o(L(t)), \qquad (t \to 0).$$

Firstly, we prove that the relation

$$(5) \qquad \int_0^t \rho'(u) u^p du = o(t^p L(t))$$

holds true for every positive p, as $t \to 0$.

In fact, in virtue of $t^p \rho(t) = o(1)$, integration by parts gives

$$(6) \qquad \int_0^t \rho'(u) u^p du = \rho(t) t^p - \int_0^t \rho(u) \cdot p u^{p-1} du.$$

Using the relation $\rho(t) = o(L(t))$, we have

$$(7) \quad \int_0^t \rho(u) \cdot p u^{p-1} du = o\left(\int_0^t L(u) \cdot p u^{p-1} du\right) = o\left(L(t) t^p + \int_0^t u^{p-1} \rho(u)\,du\right),$$

in view of the fact that $L(t) t^p = t^p \int_t^\pi o(u^{-1-p/2})\,du = o(1)$.

Since $\rho(u) > 0$, it follows from (7) that

$$\int_0^t \rho(u) \cdot p u^{p-1} du = o(L(t) t^p).$$

On combining this relation with (6), we obtain (5), by using $\rho(t) = o(L(t))$ again.

On integration by parts, we obtain

$$(1/t) \int_0^t \rho(u)\,du = \rho(t) - (1/t) \int_0^t u \rho'(u)\,du = o(L(t)),$$

by appealing to (5) and (3). This establishes the lemma.

7. Lemma 4. *If the Fourier series of $f(t)$, at $t = x$, is summable $l \mid C, \alpha \mid$ $(\alpha > -1)$ to the sum 0, then there is a sequence $\{t_n\}$ such that*

$$0 < t_n \to 0, \qquad \lim \Psi(t_n) = 0,$$

where $l = \{l_n\}$, $l_n = L(1/n)$, $L(t) = \int_t^\pi \rho(u) u^{-1} du$, and the functions $\rho(t)$ and $\Psi(t)$ satisfy the conditions of Theorem 2.

In fact, by hypothesis, the positive function

$$\epsilon(t) = \rho(t)/L(t)$$

is $o(1)$, as $t \to 0$. We shall use the function

$$\rho = \rho(t) = \max\left(\tfrac{3}{4}, 1 + 1/2\epsilon(t)\log t\right), \qquad (0 < t < 1). \tag{8}$$

In virtue of the monotony of $\rho(t)$, the derivative $\rho'(t)$ has a definite sign for small t. If $\rho'(t) \leqq 0$, we have

$$\begin{aligned} L(t^p)/\rho(t) &= 1/\epsilon(t) - (1/\rho(t)) \int_t^{t^p} (\rho(u)/u)\,du \\ &> 1/\epsilon(t) - 1/\rho(t)\cdot\rho(t)\log t^p/t \geqq 1/2\epsilon(t), \end{aligned}$$

by (8). Or, if t is small enough,

$$\rho(t)/L(t^p) < 2\epsilon(t).$$

This result is also true, when $\rho'(t) > 0$. Indeed, if $L(0)$ is a convergent integral, then it follows from

$$\lim_{t\to 0} L(t)/L(t^p) = 1$$

that

$$\rho(t)/L(t^p) = \epsilon(t)L(t)/L(t^p) < 2\epsilon(t), \tag{9}$$

for small t. On the contrary, if $L(0) = \infty$, we have

$$\begin{aligned} 1 \leqq \limsup_{t\to 0} L(t)/L(t^p) &= \limsup_{t\to 0}\,[dL(t)/dt]/[dL(t^p)/dt] \\ &= \limsup_{t\to 0}\, \rho(t)/p\rho(t^p) \leqq 1/p \leqq 4/3. \end{aligned}$$

From this (9) follows, if t is sufficiently small.

Since the summability $l\,|\,C, \alpha\,|$ of ΣA_n implies the convergence of $\Sigma\,|\,\sigma_n^\alpha - \sigma_{n-1}^\alpha\,|$, we have by Lemma 2,

$$\begin{aligned} \Phi(t) \equiv \Psi(t)/\rho(t) &= t^\lambda[t^{-\lambda}\phi(t)]_{1+\alpha} \\ &= \Big(\sum_{n\leqq t^{-1}} + \sum_{n>t^{-1}}\Big)(\alpha+1)(\sigma_n^\alpha - \sigma_{n-1}^\alpha)F_n(t) = \Phi_1(t) + \Phi_2(t). \end{aligned}$$

By Lemma 1, we have $F_n(t) = O(1)$. Hence

$$\Phi_2(t) = O\Big(\sum_{n>t^{-1}} |\,\sigma_n^\alpha - \sigma_{n-1}^\alpha\,|\Big) = O\Big([1/L(t)]\sum_{n>t^{-1}}^{\infty} l_n\,|\,\sigma_n^\alpha - \sigma_{n-1}^\alpha\,|\Big).$$

It follows that

$$\rho(t)\Phi_2(t) = o(1), \tag{10}$$

as $t \to 0$, since $\rho(t) = o(L(t))$ and ΣA_n is summable $l\,|\,C, \alpha\,|$.

Writing $[t^{-1}] = T$, we have to consider $\Phi_1(t)$ wherein

$$F_n(t) = B(1-\lambda, 1+\alpha) + O(\nu t)^{\alpha+1},$$

by Lemma 1. Consequently, we may write

$$\Phi_1(t) = (\alpha+1)B(1-\lambda, 1+\alpha)\sigma_T^\alpha + \sum_{n=0}^{T} (\sigma_n^\alpha - \sigma_{n-1}^\alpha) O(nt)^{\alpha+1}. \tag{11}$$

Remembering that $\sigma_n^\alpha = o(1)$, we have

$$\sigma_T^\alpha = \sum_{n=T+1}^{\infty} (\sigma_{n-1}^\alpha - \sigma_n^\alpha).$$

Accordingly, $| \rho(t)\sigma_T^\alpha | \leqq \epsilon(t) \sum_{n>T} l_n | \sigma_n^\alpha - \sigma_{n-1}^\alpha |$. Hence we obtain

$$\rho(t)\sigma_T^\alpha = o(1). \tag{12}$$

In virtue of the relations $\alpha + 1 > 0$, $T\nu \leqq 1$, we have, on writing $M = M(t)[t^{-p}]$,

$$\sum_{n=0}^{T} | \sigma_n^\alpha - \sigma_{n-1}^\alpha | (nt)^{\alpha+1} \leqq (Mt)^{\alpha+1} \sum_{n=0}^{M} | \sigma_n^\alpha - \sigma_{n-1}^\alpha | + \sum_{n=M+1}^{T} | \sigma_n^\alpha - \sigma_{n-1}^\alpha |. \tag{13}$$

It follows from (8) that

$$t^{(1-p)(1+\alpha)} = \max\{t^{(1+\alpha)/4}, \exp(-(1+\alpha)/2\epsilon(t))\},$$

so that

$$\begin{aligned} T_1(t) &\leqq t^{(1-p)(1+\alpha)} \sum | \sigma_n^\alpha - \sigma_{n-1}^\alpha | \\ &= O(t^{(1+\alpha)/4}) + O(\exp(-(1+\alpha)/2\epsilon(t))). \end{aligned} \tag{14}$$

On differentiating the equation $\rho(t)/\epsilon(t) = L(t)$ with respect to t, and multiplying by $\epsilon(t)/\rho(t)$, and then integrating the equality obtained from t to 1, we get

$$\rho(t)/\epsilon(t) = C \exp(\int_t^1 (\epsilon(u)/u) du$$

with $C = L(1)$. Take a positive number K small enough such that

$$K\rho(t) \leqq | \log t |^A, \tag{15}$$

and set $G = A + 1$. If there exists a positive number t_0 such that

$$\epsilon(t) \log 1/t \geqq G$$

for $0 < t \leqq t_0$, then we have

$$\rho(t)/\epsilon(t) \geqq C \exp(\int_t^{t_0} (\epsilon(u)/u)\,du$$
$$\geqq C \exp(G \log\ (\log t/\log t_0) = C(\log t/\log t_0)^G.$$

On combining this result with (15), we obtain

$$1/\epsilon(t) \geqq (KC/|\log t_0|^G)\ |\log t| \qquad (0 < t \leqq t_0). \tag{16}$$

Writing

$$\gamma = \min((1+\alpha)/4,\ (1+\alpha)KC/2\,|\log t_0|^G),$$

it follows from (14) and (16) that

$$\begin{aligned}\rho(t)T_1(t) &= O(\rho(t)t^\gamma) + O(\exp(-\gamma\,|\log t|)) \\ &= O(t^\gamma\,|\log t|^A) = o(1).\end{aligned} \tag{17}$$

On the contrary, if there exists no t_0 such as assumed above, then there is a sequence of positive numbers t_ν, $t_\nu \to 0$, satisfying

$$\epsilon(t_\nu) \log 1/t_\nu < G;$$

and (14) gives

$$\begin{aligned}\rho(t_\nu)T_1(t_\nu) &= O(\rho(t_\nu)t^{(\alpha+1)/2}) + O(\rho(t_\nu))\exp(-(1+\alpha)/2G \log 1/t_\nu) \\ &= O(t_\nu{}^{(\alpha+1)/2G}\,|\log t_\nu|^A) = o(1).\end{aligned} \tag{18}$$

By (17), the relation (18) holds good for any case.

Finally, it follows from

$$T_2(t) = \sum_{M+1}^{N} |\sigma_n{}^\alpha - \sigma_{n-1}{}^\alpha| \leqq (1/L(t^p)) \sum_{M+1}^{N} l_n\,|\sigma_n{}^\alpha - \sigma_{n-1}{}^\alpha|$$

and $\rho(t) < 2L(t^p)\epsilon(t)$ that

$$\rho(t)T_2(t) < 2\epsilon(t) \sum l_n\,|\sigma_n{}^\alpha - \sigma_{n-1}{}^\alpha| = o(1). \tag{19}$$

Combining the results (10), (11), (12), (13), (18) and (19), we arrive at the relation

$$\Psi(t_\nu) = \rho(t_\nu)\Phi(t_\nu) = \rho(t_\nu)\Phi_1(t_\nu) + \rho(t_\nu)\Phi_2(t_\nu) = o(1).$$

This proves the lemma.

8. We can now prove Theorem 2. It follows from Lemmas 1 and 2 that the relation

$$(20)\qquad \Phi'(t) = (1+\alpha)\sum_{n=0}^{\infty}(\sigma_n{}^{\alpha} - \sigma_{n-1}{}^{\alpha})F'_n(t)$$

holds true for $0 < t \leqq \pi$, since $F'_n(t) = O(t^{-1})$ and the series

$$\sum |\sigma_n{}^{\alpha} - \sigma_{n-1}{}^{\alpha}|$$

is convergent.

Using Lemma 1, we have

$$\int_0^{\pi}\rho(t)\,|F'_n(t)|\,dt = \int_0^{1/n}\rho(t)\,|F'_n(t)|\,dt + \int_{1/n}^{\pi}\rho(t)\,|F'_n(t)|\,dt$$
$$= \int_0^{1/n}\rho(t)O(n)\,dt + \int_{1/n}^{\pi}\rho(t)t^{-1}dt.$$

The second term is l_n. And the first term is by Lemma 3, $o(l_n)$ as $n \to \infty$. Hence

$$(21)\qquad \int_0^{\pi}\rho(t)\,|F'_n(t)|\,dt \sim l_n.$$

By (20) and (21), there is a constant K satisfying

$$(22)\qquad \int_0^{\pi}|\rho(t)\Phi'(t)|\,dt \leqq K\sum_{n=0}^{\infty} l_n\,|\sigma_n{}^{\alpha} - \sigma_{n-1}{}^{\alpha}|.$$

Now, let t_n be the numbers given by Lemma 4; then $\Psi(t_n) = o(1)$. Integration by parts gives

$$\int_0^{\pi}\rho'(t)\,|\Phi(t)|\,dt = \lim_{n\to\infty}\int_{t_n}^{\pi}\rho'(t)\,|\Phi(t)|\,dt$$
$$= \rho(\pi)\,|\Phi(\pi)| - \int_0^{\pi}\rho(t)\,(d/dt)\,|\Phi(t)|\,dt.$$

In virtue of the monotony of $\rho(t)$, the above relation implies that

$$\int_0^{\pi}|\rho'(t)\Phi(t)|\,dt \leqq \rho(\pi)\,|\Phi(\pi)| + \int_0^{\pi}\rho(t)\,|\Phi'(t)|\,dt.$$

Hence, by (22),

$$(23)\qquad \int_0^{\pi}|\rho'(t)\Phi(t)|\,dt \leqq \rho(\pi)\,|\Phi(\pi)| + K\sum l_n\,|\sigma_n{}^{\alpha} - \sigma_{n-1}{}^{\alpha}|.$$

We have $\Psi'(t) = \rho(t)\Phi'(t) + \rho'(t)\Phi(t)$. Hence, by (22) and (23), we obtain

$$\int_0^{\pi}|\Psi'(t)|\,dt \leqq \rho(\pi)\,|\Phi(\pi)| + 2K\sum l_n\,|\sigma_n{}^{\alpha} - \sigma_{n-1}{}^{\alpha}|.$$

This establishes Theorem 2.

9. In his paper "The absolute Cesàro summability of a Fourier series,"[4] Bosanquet asserts nothing concerning the behavior of the function

$$[\phi(t)]_{1+\alpha} = (1+\alpha)t^{-1-\alpha}\int_0^t (t-u)^\alpha \phi(u)\,du,$$

when the Fourier series ΣA_n is summable $|C,\alpha|$ $(\alpha \geqq 0)$. It should be observed that the summability $|C,\alpha|$, is equivalent to the summability

$$\{L(1/n)\}\,|C,\alpha| \qquad (\alpha > -1)$$

in the case where $L(t)$ converges to a finite number, as $t \to 0$. Accordingly we can infer from Theorem 2 the following proposition.

THEOREM 3. *If the Fourier series of $f(t)$, at $t = x$, is summable $|C,\alpha|$ $(\alpha > -1)$, then the function*

$$\rho(t)t^\lambda[t^{-\lambda}\phi(t)]_{1+\alpha} \qquad (\lambda < 1,\ \lambda \leqq 1-(\alpha))$$

is of bounded variation in $(0,\pi)$, where $\rho(t)$ is any function which is absolutely continuous in $(0,\pi)$ such that

$$\int_0^\pi |t^{-1}\rho(t)|\,dt < \infty,$$

and where $2\phi(t)$ denotes $f(x+t)+f(x-t)$.

In fact, the relation (21) is now reduced to

$$\int_0^\pi |\rho(t)F'_n(t)|\,dt = O(1).$$

Consequently, (22) becomes

$$\int_0^\pi |\rho(t)\Phi'(t)|\,dt \leqq K\,|\sigma_n^\alpha - \sigma_{n-1}^\alpha|. \tag{24}$$

We next have

$$\int_0^\pi |\rho'(t)F_n(t)|\,dt = O(1),$$

by the relations $F_n(t) = O(1)$ and $\int_0^\pi |\rho'(t)|\,dt < \infty$. It follows that

$$\int_0^\pi |\rho'(t)\Phi(t)|\,dt \leqq K'\sum |\sigma_n^\alpha - \sigma_{n-1}^\alpha|, \tag{25}$$

K' being a constant.

---

[4] L. S. Bosanquet (2).

On combining the results (24) and (25), we obtain

$$\int_0^\pi |(d/dt)\{\rho(t)\Phi(t)\}|\,dt \leqq (K+K')\sum |\sigma_n{}^\alpha - \sigma_{n-1}{}^\alpha|.$$

The theorem is thus proved.

NATIONAL UNIVERSITY OF CHEKIANG,
MEITAN, KWEICHOW, CHINA.

REFERENCES.

L. S. Bosanquet.

(1) "Note on the absolute summability (C) of a Fourier series," *Journal of the London Mathematical Society*, vol. 11 (1936), pp. 11-15.

(2) "The absolute Cesàro summability of a Fourier series," *Proceedings of the London Mathematical Society* (2), vol. 41 (1935), pp. 517-528.

K. K. Chen.

(1) "The super-absolute Cesàro summability of Fourier series," *Science Record, Academia Sinica*, vol. 1.

A GENERALIZATION OF HARDY'S THEOREM WITH AN APPLICATION TO THE ABSOLUTE SUMMABILITY OF FOURIER SERIES.*

By KIEN-KWONG CHEN.

1. Suppose that $f(x)$ is integrable in the Lebesgue sense on any finite interval, and that $f(x+2\pi)=f(x)$. Then $[\phi(t)]_\alpha$, the mean function of f, at the point x, of order α, may be defined by the equation

$$[\phi(t)]_\alpha=\alpha t^{-\alpha}\int_0^t (t-u)^{\alpha-1}\phi(u)\,du$$

where $\alpha>0$, $2\phi(t)=f(x+t)+f(x-t)$. If $[\phi(t)]_1$ is of bounded variation in $(0,\pi)$, then x is said to be a point of de la Vallée-Poussin for f. At such a point, the Fourier series of f is summable $|C,\alpha|$ for every $\alpha>1$.[1] The point x is said to be a Dini point of f, if $t^{-1}\phi(t)$ is integrable in the Lebesgue sense on $(0,\pi)$.

Hardy[2] has proved that a Dini point is a point of de la Vallée-Poussin. Evidently, the integrability of $t^{-1}\phi(t)$ depends only on the behavior of the function f in the neighborhood of the point x. Bosanquet and Kestelman[3] have shown that the summability $|C,1|$ for a Fourier series at a given point is not a local property of the function under consideration. Therefore, at a Dini point of f, the Fourier series is not necessarily summable $|C,1|$.

Theorem 1 and Theorem 4 of this paper are generalizations of Hardy's theorem. From these propositions, we derive Theorem 2 and Theorem 3. Direct proofs of the latter theorems are also given.

2. We begin with the following

LEMMA 1. *If $l(t)\in L(0,\pi)$, then*

$$h(t)=t^{-3}\int_0^t\int_0^u vl(v)\,dv\,du\in L(0,\pi).$$

To prove this proposition, we may assume that $l(t)$ is non-negative. Then we have

---

* Received August 17, 1943.

[1] L. S. Bosanquet 1.

[2] G. H. Hardy 1.

[3] L. S. Bosanquet and H. Kestelman 1.

$$h(t) = (1/t^2) \int_0^t wl(w)\,dw - \int_0^t w^2 l(w)\,dw \leqq (1/t^2) \int_0^t wl(w)\,dw$$

and the result is a consequence of the following lemma:

LEMMA 2. *If* $l(t) \in L(0, \pi)$, *then*

$$t^{-2} \int_0^t ul(u)\,du \in L(0, \pi).$$

For the proof, we can assume that $l(t) \geqq 0$. It follows from Hardy's theorem that the function

$$L(t) = (1/t) \int_0^t ul(u)\,du$$

is of bounded variation in $(0, \pi)$. This involves the existence of the limit $L(+0)$. Integrating by parts,

$$\int_0^\pi l(t)\,dt = L(\pi) - L(0) + \int_0^\pi t^{-2} \int_0^t ul(u)\,du\,dt.$$

This establishes Lemma 2.

3. *If the Fourier series of f is absolutely summable C at the point x, then* $t^{-\lambda}\phi(t) \in L(0, \pi)$ *for every* $\lambda < 1$, *but not for* $\lambda = 1$. The proof of this theorem will be given in another paper.

Let us now assume that the integral

$$\chi(t) = \int_{+0}^{\pi} \phi(u) u^{-1}\,du$$

exists in the Cauchy sense, i. e. the limit $\lim\limits_{\epsilon \to +0} \int_\epsilon^\pi$ exists as a finite number. Assume further that the function $l(t) = t^{-1}\chi(t)$ is integrable in the Lebesgue sense on $(0, \pi)$. Then, the mean function $[\phi(t)]_2$ of the second order is of bounded variation in $(0, \pi)$.

In fact we have

$$[\phi(t)]_2 \doteq (2/t^2) \int_0^t \int_0^u v \cdot \phi(v) v^{-1}\,dv\,du$$

$$= (2/t^2) \int_0^t u\chi(u)\,du - (2/t^2) \int_0^t \int_0^u \chi(v)\,dv\,du.$$

Integration by parts yields

$$\int_0^t u\chi(u)\,du = t \int_0^t \chi(u)\,du - \int_0^t \int_0^u \chi(v)\,dv\,du.$$

On differentiation, we obtain

$$(d/dt)[\phi(t)]_2 = (2/t)\chi(t) - (6/t^2)\int_0^t \chi(u)\,du + (8/t^3)\int_0^t \int_0^u \chi(v)\,dv\,du.$$

Hence, by Lemma 1 and Lemma 2, we have

$$\int_0^\pi |(d/dt)[\phi(t)]_2|\,dt < \infty.$$

The following theorem has thus been established:

THEOREM 1. *If* $t^{-1}\chi(t) \in L(0,\pi)$, *then* $[\phi(t)]_2$ *is of bounded variation in* $(0,\pi)$.

A combination of Theorem 1 and a Theorem of Bosanquet[4] gives

THEOREM 2. *If* $t^{-1}\chi(t) \in L(0,\pi)$, *then the Fourier series of* f *at the point* x *is summable* $|C, \alpha|$ *for every* $\alpha > 2$.

4. Next, we shall give a direct proof of Theorem 2. This is based on the following two lemmas.

LEMMA 3. *Suppose that* $a < 1$, $b > 1$ *and* $h(t) \geqq 0$. *Then the existence of the integral* $\int_0^t h(t)t^{-1}dt$ *implies the convergence of the following two series*

$$\sum_1^\infty n^{-a}\int_0^{1/n} t^{-a}h(t)\,dt \quad \textit{and} \quad \sum_1^\infty n^{-b}\int_{1/n}^1 t^{-b}h(t)\,dt.$$

In fact, we have

$$u_n = n^{-a} {}_0^{1/n} t^{-a}h(t)\,dt = \int_0^{1/n} \min\{(nt)^{-a}, (nt)^{-b}\}h(t)\,dt$$

$$v_n = n^{-b}\int_{1/n}^1 t^{-b}h(t)\,dt = \int_{1/n}^1 \min\{(nt)^{-a}, (nt)^{-b}\}h(t)\,dt.$$

Writing

$$H(t) = \sum_{n=1}^\infty t\min\{(nt)^{-a}, (nt)^{-b}\}$$

we have

$$\sum u_n \leqq \int_0^1 H(t)t^{-1}h(t)\,dt, \qquad \sum v_n \leqq \int_0^1 H(t)t^{-1}h(t)\,dt.$$

On account of the hypotheses $h(t) \geqq 0$ and $t^{-1}h(t) \in L(0,1)$, the convergence of the two series Σu_n and Σv_n follows from the boundedness of $H(t)$. We have

---

[4] L. S. Bosanquet 1.

$$H(t) = t^{1-a} \sum_{nt \leqq 1} n^{-a} + t^{1-b} \sum_{nt > 1} n^{-b}$$
$$< \frac{t^{1-a}}{1-a} (1/t)^{1-a} + \frac{t^{1-b}}{b-1} (1/t)^{1-b} + 2 = \frac{2-a}{1-a} + \frac{b}{b-1}.$$

This completes the proof of the lemma.

LEMMA 4. *Write $g^{\alpha}(n, t)$ for the n-th Cesàro mean of order α of $2\pi^{-1} \sin nt$ $(n = 0, 1, 2, \cdots)$; then for $\alpha > 0$, $0 < t < \pi$, $n > 0$, $\lambda \geqq 0$*

$$|(d/dt)^{\lambda} g^{\alpha}(n, t)| \leqq K n^{\lambda} (1 + nt)^{-\mu}$$

where $\mu = \min(\alpha, 1 + \lambda)$, K being a constant.

This is a known theorem.[5]

Let $2 < \beta < 3$, and write r_n^{β} and σ_n^{β} for the n-th Cesàro mean of order β of the sequences

$$\{\tfrac{1}{2}a_0 + a_1 + \cdots + a_n\} \quad \text{and} \quad \{na_n\}$$

respectively, where a_n $(n = 0, 1, \cdots)$ denote the Fourier coefficients of the function $\phi(t)$. If the series

$$\sum_{n=0}^{\infty} (\sigma_n^{\beta} - \sigma_{n-1}^{\beta})$$

with $\sigma_{-1}^{\beta} = 0$, converges absolutely to the sum S, then we write

$$\tfrac{1}{2}a_0 + \sum_{1}^{\infty} a_n = s \mid C, \beta \mid.$$

An elementary calculation gives

$$\sigma_n^{\beta} - \sigma_{n-1}^{\beta} = n^{-1} r_n^{\beta}.$$

We have to show that the convergence of $\sum n^{-1} \mid r_n^{\beta} \mid$ follows from

$$t^{-1}\chi(t) \in L(0, \pi).$$

Since

$$na_n = (2/\pi) \int_0^{\pi} \phi(t) (d/dt) \sin nt dt,$$

we have

$$r_n^{\beta} = (2/\pi) \int_0^{\pi} \phi(t) (d/dt) g^{\beta}(n, t) dt.$$

Now, on integration by parts,

---

[5] N. Obrechkoff 1. Cf. also L. S. Bosanquet 1, p. 519.

$$\int_0^\pi \phi(t)\,(d/dt)\,g^\beta(n,t)\,dt$$
$$= \left[t\chi(t)\,(d/dt)\,g^\beta(n,t)\right]_0^\pi - \int_0^\pi \chi(t)\,(d/dt)\{t(d/dt)\,g^\beta(n,t)\}dt.$$

This is equal to

$$O(n^{-1}) + \int_0^\pi \chi(t)\,O(1+nt)^{-2}dt + n^2\int_0^\pi t\chi(t)\,O(1+nt)^{-\beta}dt,$$

in view of Lemma 4. Hence there is a constant K such that

$$\Sigma\,|\,r_n{}^\beta/n\,| \leqq K\Sigma(1/n^2) + K\Sigma\int_0^{1/n}|\,\chi(t)\,|\,dt + K\Sigma n^{-2}\int_{1/n}^\pi t^{-2}\,|\,\chi(t)\,|\,dt$$
$$+ K\Sigma n\int_0^{1/n} t\,|\,\chi(t)\,|\,dt + K\Sigma n^{1-\beta}\int_{1/n}^\pi t^{1-\beta}\,|\,\chi(t)\,|\,dt.$$

The convergence of the last four series is, by Lemma 3, a consequence of the integrability of $t^{-1}\,|\,\chi(t)\,|$. This establishes the relation

$$\tfrac{1}{2}a_0 + \sum_1^\infty a_n = S\,|\,C,\beta\,|,$$

for $2 < \beta < 3$. The proof of Theorem 2 may be completed by the following theorem of Kogbetliantz.[6]

LEMMA 5. *If* $\alpha > 0$, *then* $\Sigma a_n = S\,|\,C,\alpha\,|$ *implies* $\Sigma a_n = S\,|\,C,\beta\,|$ *for* $\beta > \alpha$.

5. Suppose that $l(t) \,\epsilon\, L(0,\pi)$. Write

$$h_n(t) = t^{-1-n}\int_0^t\int_0^{t_{n-1}}\int_0^{t_{n-2}}\cdots\int_0^{t_1} t_0 l(t_0)\,dt_0dt_1\cdots dt_{n-1}.$$

Lemma 1 can be generalized as follows:

LEMMA 6. *The relation* $l(t) \,\epsilon\, L(0,\pi)$ *implies* $h_n(t) \,\epsilon\, L(0,\pi)$, *for every positive integer.*

In fact, integration by parts gives

$$h_n(t) = h_{n-1}(t) - t^{-1-n}\int_0^t u^n h_{n-2}(u)\,du.$$

To obtain the relations $h_n(t) \,\epsilon\, L(0,\pi)$ $(n = 1, 2, \cdots)$, we can assume that

---

[6] E. Kogbetliantz 1. See also E. Kogbetliantz 2.

$l(t) \geqq 0$ so that $h_n(t) \leqq h_{n-1}(t)$, for $n = 2, 3, \cdots$. Then, Lemma 2 completes the proof. The arguments developed in the preceding article can be used to extend Theorem 2.

THEOREM 3. *If the repeated Cauchy integral*

$$l(t) \equiv (1/t) \int_{+0}^{t} (1/t_{m-1}) \int_{+0}^{t_{m-1}} (1/t_{m-2}) \int_{+0}^{t_{m-2}} \cdots (1/t_1) \int_{+0}^{t_1} (\phi(t_0)/t_0)\, dt_0 dt_1 \cdots dt_n$$

exists as a function integrable in the Lebesgue sense on the interval $(0, \pi)$, *then the Fourier series of* f, *at the point* x, *is absolutely summable* $(C, m + 1 + \epsilon)$ *for every positive* ϵ.

In fact, let $\beta > m + 1$; using Lemma 4, integration by parts gives

$$\begin{aligned} r_n^{\beta} &= (-1)^m \int_0^{\pi} l(t)\,(t(d/dt))^{m+1} g^{\beta}(n, t)\, dt + O(1/n) \\ &= \int_0^{\pi} l(t) \sum_{k=0}^{m} (1 + nt)^{-1-k} O(nt)^k dt \\ &\quad + \int_0^{\pi} l(t)\,(1 + nt)^{-\gamma} O(nt)^{m+1} dt + O(1/n), \end{aligned}$$

where $\gamma = \min(\beta + m + 2)$. Hence, there is a constant K such that

$$\begin{aligned} | r_n^{\beta}/n | \leqq K \sum_{k=0}^{m+1} n^{k-1} \int_0^{1/n} t^{k-1} \,| tl(t) |\, dt + K n^{-2} \int_{1/n}^{\pi} t^{-2} \,| tl(t) |\, dt \\ + K n^{m-\gamma} \int_{1/n}^{\pi} t^{m-\gamma} \,| tl(t) |\, dt + O(1/n^2). \end{aligned}$$

From this, the convergence of $\Sigma n^{-1} \,| r_n^{\beta} |$ follows, in view of Lemma 3.

Theorem 3 can also be deduced from Bosanquet's criterion. Indeed, we have the following theorem:

THEOREM 4. *If the repeated Cauchy integral* $l(t)$ *in Theorem 3 exists as a function integrable in the Lebesgue sense on* $(0, \pi)$, *then the mean function* $[\phi(t)]_{m+1}$ *of order* $1 + m$ *is of bounded variation in* $(0, \pi)$.

Write $\phi(t) = \phi_0(t) \equiv \phi_0$, and for $\nu > 0$,

$$\phi_\nu \equiv \phi_\nu(t) = \int_{+0}^{t} \phi_{\nu-1}(t)\, t^{-1} dt;$$

then $l(t) = t^{-1} \phi_m(t)$. From

$$\begin{aligned} t^{m+1} [\phi(t)]_{m+1}/(m + 1) = (\int_0^t dt)^{m+1} \phi(t) &= (\int_0^t dt)^m \{ t\phi_1 - \int_0^t \phi_1 dt \} \\ &= (\int_0^t dt)^m t\phi_1 - (\int_0^t dt)^{m+1} \phi_1, \end{aligned}$$

we infer that

$$t^{m+1}[\phi(t)]_{m+1}/(m+1) = t\Big(\int_0^t dt\Big)^m \phi_1 - (m+1)\Big(\int_0^t dt\Big)^{m+1}\phi_1.$$

so that $t^{m+1}[\phi(t)]_{m+1}$ is a linear combination of the functions

$$t^m\Big(\int_0^t dt\Big)^1\phi_m,\ t^{m-1}\Big(\int_0^t dt\Big)^2\phi_m, \cdots, \Big(\int_0^t dt\Big)^{m+1}\phi_m.$$

Consequently, the function $[\phi(t)]_{m+1}$ is a linear combination of the functions

$$t^{-1}\int_0^t tl(t)\,dt,\ t^{-2}\Big(\int_0^t dt\Big)^2 tl(t), \cdots, t^{-1-m}\Big(\int_0^t dt\Big)^{1+m} tl(t);$$

or of the functions

$$th_1(t), th_2(t), \cdots, th_{m+1}(t).$$

On differentiation, we find that

$$(d/dt)[\phi(t)]_{m+1} = c_0 l(t) + c_1 h_1(t) + \cdots + c_{m+1} h_{m+1}(t),$$

where $c_0, c_1, \cdots, c_{m+1}$ denote constants. Since $l(t)$, and by Lemma 6, $h_\nu(t)$ are integrable in the Lebesgue sense on $(0, \pi)$, it follows that $[\phi(t)]'_{m+1}$ is absolutely integrable on $(0, \pi)$. Hence $[\phi(t)]_{m+1}$ is of bounded variation in $(0, \pi)$.

NATIONAL UNIVERSITY OF CHEKIANG,
MEITAN, KWEICHOW, CHINA.

REFERENCES.

L. S. Bosanquet 1, "The absolute Cesàro summability of a Fourier series," *Proceedings of the London Mathematical Society* (2), vol. 41 (1935), pp. 517-528.

L. S. Bosanquet and H. Kestelman 1, "The absolute convergence of series of integrals," *Proceedings of the London Mathematical Society* (2), vol. 45 (1930), pp. 88-97.

G. H. Hardy 1, "Note on some points in the integral calculus LIII: on certain criteria for the convergence of the Fourier series of a continuous function," *Messenger of Mathematics*, vol. 49 (1920), pp. 49-156.

E. Kogbetliantz 1, "Sur les séries absolument sommables par la méthode des moyennes arithmétiques," *Bull. des Sciences Math.* (2), vol. 49 (1925), pp. 234-256.

E. Kogbetliantz 2, "Sommation des séries et integrales divergentes par les moyennes arithmétiques et typiques," *Mémorial des Sciences Mathématiques* LI (1931), Théorème VIII, p. 25.

N. Obrechkoff 1, "Sur la sommation des séries trigonométriques de Fourier par les moyennes arithmétiques," *Bull. de la Soc. Math. de France*, vol. 62 (1934), pp. 84-109 and pp. 167-184.

W. C. Randels 1, "On the absolute summability of Fourier series," *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 733-736.

THE LAGRANGE MULTIPLIER RULE FOR TWO DEPENDENT AND TWO INDEPENDENT VARIABLES.*

By CHARLES B. BARKER, JR.

Introduction. The Lagrange Multiplier Rule in the Calculus of Variations may be briefly described as follows. Suppose that $\bar{z}_1(x,y)$ and $\bar{z}_2(x,y)$ are functions which minimize an integral of the form

$$\iint_G f(x, y, z_1, z_2, z_{1x}, z_{1y}, z_{2x}, z_{2y})\,dydx$$

among all functions $z_1(x,y)$ and $z_2(x,y)$ which assume certain prescribed boundary values on G^* and satisfy a given differential equation,

$$\phi(x, y, z_1, z_2, z_{1x}, z_{1y}, z_{2x}, z_{2y}) = 0.$$

Then there exists a function $\lambda(x,y)$, which is continuous together with its first derivative on the closed region $\bar{G}$ and is such that it satisfies the equations of variation,

$$(\partial/\partial x)(f_{z_{1x}} + \lambda\phi_{z_{1x}}) + (\partial/\partial y)(f_{z_{1y}} + \lambda\phi_{z_{1y}}) = f_{z_1} + \lambda\phi_{z_1}$$
$$(\partial/\partial x)(f_{z_{2x}} + \lambda\phi_{z_{2x}}) + (\partial/\partial y)(f_{z_{2y}} + \lambda\phi_{z_{2y}}) = f_{z_2} + \lambda\phi_{z_2}$$

in which the unwritten arguments of the partial derivatives of f and ϕ are $\bar{z}_1(x,y)$ and $\bar{z}_2(x,y)$ and their partial derivatives.

The existence of the function $\lambda(x,y)$ is demonstrated in this paper for all $\bar{z}_1(x,y)$ and $\bar{z}_2(x,y)$ which form a 'quasi-normal surface,' an explicit definition of which will be given later. The first result obtained in a study of the Lagrange problem for double integrals was that of Gross [4].[1] His result was re-established by Coral [3] under weakened conditions. A further result obtained by Coral in the same paper is closely related to the result of this paper. However, the region of integration considered by Coral is considerably more restricted than the one considered here, and Coral's differential equation does not contain all of the partial derivatives of the functions. On the other hand, the condition of 'quasi-normality' imposed upon the surfaces considered in the present paper appears to be more restrictive than the

---

* Received July 11, 1944.

[1] Numbers in brackets refer to the bibliography.

'normal' condition of Coral's. A very comprehensive treatment of the problem in one dimension, together with an extensive bibliography may be found in [2].

1. Notation. The following notation is employed. If G is a region, then $\bar{G}$ indicates its closure and G^* its boundary. A function, $f(x,y)$, is said to be of class $C^{(n)}$ on $\bar{G}$ if the function, together with its partial derivatives of the first n orders, are continuous on G and coincide with functions which are continuous on $\bar{G}$. A region G is said to be bounded by a simple, closed, regular curve of class $C_\alpha^{(n)}$ if G^* is rectifiable and if the parametric equations of G^*, $x = x(s)$, $y = y(s)$, s being the arc length, are of class $C^{(n)}$ with $x^{(n)}(s)$ and $y^{(n)}(s)$ satisfying a uniform Hölder condition with exponent α, $0 < \alpha < 1$. The symbols p_i and q_i are used to represent, respectively,

$$(\partial/\partial x)[z_i(x,y)] \quad \text{and} \quad (\partial/\partial y)[z_i(x,y)], \qquad (i = 1, 2).$$

2. A preliminary transformation. In this section, a particular one-to-one transformation of the entire (x,y)-plane into the entire (ξ,η)-plane is established. This mapping possesses specific properties which are essential to the solution of the problem. In order to obtain the desired transformation, it is necessary to prove certain preliminary lemmas.

LEMMA 1. *Let G be a region bounded by a simple, closed, regular curve of class C_α'''. Suppose that $f(x,y)$ and $g(x,y)$ are functions of class C''' on $\bar{G}$, and that*

$$[f(x,y)]^2 + [g(x,y)]^2 > 0 \tag{1}$$

at each point of $\bar{G}$. Then there exist two functions, $r(x,y)$ and $\theta(x,y)$ which are of class C''' on $\bar{G}$ and which satisfy

$$r(x,y)\cos\theta(x,y) = f(x,y), \tag{2}$$

$$r(x,y)\sin\theta(x,y) = g(x,y) \tag{3}$$

at each point of $\bar{G}$.

Proof. Let

$$r(x,y) = +\{[f(x,y)]^2 + [g(x,y)]^2\}^{\frac{1}{2}}, \tag{4}$$

and let

$$\theta(x,y) = \int_{(x_0,y_0)}^{(x,y)} [A(x,y)\,dx + B(x,y)\,dy] + \theta(x_0,y_0) \tag{5}$$

where (x_0, y_0) is a point of $\bar{G}$, and where

$$A(x,y) = \frac{fg_x - f_x g}{f^2 + g^2}, \qquad B(x,y) = \frac{fg_y - f_y g}{f^2 + g^2}. \tag{6}$$

By Green's Theorem, if R is any rectangle contained in $\bar{G}$,

$$\int_{R^*} [A(x,y)\,dx + B(x,y)\,dy] = 0.$$

Hence, by a well known theorem (see, for example, [6]), $\theta(x,y)$ is of class C' on $\bar{G}$, and is such that

$$\partial\theta/\partial x = A(x,y), \qquad \partial\theta/\partial y = B(x,y). \tag{7}$$

Further, since

$$A(x,y) = (\partial/\partial x)(\text{arc} \tan g/f), \qquad B(x,y) = (\partial/\partial y)(\text{arc} \tan g/f),$$

it follows that

$$\cos\theta = f/\pm\{f^2 + g^2\}^{\frac{1}{2}}, \qquad \sin\theta = g/\pm\{f^2 + g^2\}^{\frac{1}{2}}, \tag{8}$$

where the signs in the denominators are to be taken the same. Thus, if $\theta(x_0, y_0)$ in (5) is chosen so that the positive sign holds in the first place, that sign will always hold on $\bar{G}$, since $\theta(x,y)$ is continuous there. In view of (1), (4), (6), (7), and (8), $r(x,y)$ and $\theta(x,y)$ are of class C''' on $\bar{G}$, and they clearly satisfy (2) and (3).

LEMMA 2. *Let G be a region bounded by a simple, closed, regular curve of class C_a''', and let $a(x,y)$, $b(x,y)$, $c(x,y)$, and $d(x,y)$ be functions of class C''' on $\bar{G}$, with*

$$[a(x,y)][d(x,y)] - [b(x,y)][c(x,y)] \neq 0 \tag{9}$$

at each point of $\bar{G}$. Then there exist four functions, $r_1(x,y)$, $\theta_1(x,y)$, $r_2(x,y)$, and $\theta_2(x,y)$ which are of class C''' on $\bar{G}$, and are such that

$$\begin{aligned} r_1(x,y)\cos\theta_1(x,y) &= a(x,y)\,; & r_1(x,y)\sin\theta_1(x,y) &= c(x,y)\,; \\ r_2(x,y)\cos\theta_2(x,y) &= b(x,y)\,; & r_2(x,y)\sin\theta_2(x,y) &= d(x,y), \end{aligned} \tag{10}$$

on $\bar{G}$. In addition, $\theta_1(x,y)$ and $\theta_2(x,y)$ satisfy the inequality

$$0 < m \leq |\theta_2(x,y) - \theta_1(x,y)| \leq M < \pi, \tag{11}$$

at each point of $\bar{G}$.

Proof. Statement (9) implies that

$$[a(x,y)]^2 + [c(x,y)]^2 \neq 0, \quad \text{and} \quad [b(x,y)]^2 + [d(x,y)]^2 \neq 0$$

on $\bar{G}$. Thus Lemma 1 applies, and hence there exist four functions $\bar{r}_1(x,y)$, $\bar{\theta}_1(x,y)$, $\bar{r}_2(x,y)$, and $\bar{\theta}_2(x,y)$ which satisfy (10) on $\bar{G}$ and are of class C''' there. Now,

$$ad - bc = \bar{r}_1\bar{r}_2 \sin(\bar{\theta}_1 - \bar{\theta}_2),$$

and so, as a result of (9),

$$\bar{\theta}_2(x,y) - \bar{\theta}_1(x,y) \neq m\pi$$

for all integral values of m. Since both $\bar{\theta}_1(x,y)$ and $\bar{\theta}_2(x,y)$ are continuous, there exists an integer N such that

$$N\pi < [\bar{\theta}_2(x,y) - \bar{\theta}_1(x,y)] < (N+1)\pi,$$

or,

$$0 < [\{\bar{\theta}_2(x,y) - N\pi\} - \bar{\theta}_1(x,y)] < \pi.$$

Accordingly, if

$$\begin{aligned} r_1(x,y) &= \bar{r}_1(x,y), & \theta_1(x,y) &= \bar{\theta}_1(x,y), \\ r_2(x,y) &= \pm\, \bar{r}_2(x,y), & \theta_2(x,y) &= \bar{\theta}_2(x,y) - N\pi, \end{aligned}$$

where the plus or minus sign is taken according as N is even or odd, it is seen that equations (10) are satisfied and that

$$0 < |\theta_2(x,y) - \theta_1(x,y)| < \pi. \tag{12}$$

In view of the continuity of $\theta_1(x,y)$ and $\theta_2(x,y)$, (12) is equivalent to (11), thus concluding the proof of the lemma.

LEMMA 3. *Let $u_1(x)$, $u_2(x)$, $u_3(x)$, and $u_4(x)$ be functions of class C''', C'', C', and C, respectively. Then there exists a function $U(x,y)$ which is of class C''' for $y \geq 0$, and is such that*

$$\begin{aligned} U(x,0) &= u_1(x), & U_{yy}(x,0) &= u_3(x), \\ U_x(x,0) &= u_1'(x), & U_{xxx}(x,0) &= u_1'''(x), \\ U_y(x,0) &= u_2(x), & U_{xxy}(x,0) &= u_2''(x), \\ U_{xx}(x,0) &= u_1''(x), & U_{xyy}(x,0) &= u_3'(x), \\ U_{xy}(x,0) &= u_2'(x), & U_{yyy}(x,0) &= u_4(x). \end{aligned}$$

Proof. Let

$$v_1(x,y) = u_1(x)\,; \qquad v_2(x,y) = (y/1!)\left[(1/2y)\int_{x-y}^{x+y} u_2(\xi)\,d\xi\right];$$

$$v_3(x,y) = (y^2/2!)\left\{(1/2y)\int_{x-y}^{x+y}\left[(1/2y)\int_{\xi-y}^{\xi+y} u_3(\eta)\,d\eta\right]d\xi\right\};$$

$$v_4(x,y) = y^3/3!\left\{(1/2y)\times\int_{x-y}^{x+y}\left\{(1/2y)\int_{\xi-y}^{\xi+y}\left[(1/2y)\int_{\eta-y}^{\eta+y}(u_4(\sigma)-u_2''(\sigma))\,d\sigma\right]d\eta\right\}d\xi\right\}.$$

If, now, $U(x,y)$ be defined by

$$U(x,y) = v_1(x,y) + v_2(x,y) + v_3(x,y) + v_4(x,y),$$

it may easily be verified that $U(x,y)$ satisfies the conditions of the lemma.

LEMMA 4. *Let G be a region bounded by a simple, closed, regular curve of class C_a''', and let $a(x,y)$, $b(x,y)$, $c(x,y)$, $d(x,y)$, $e(x,y)$, and $f(x,y)$ be functions of class C''' on $\bar{G}$ with*

$$(13) \qquad [a(x,y)][d(x,y)] - [b(x,y)][c(x,y)] \neq 0$$

at each point of $\bar{G}$. Then these functions may be extended to be of class C''' over the entire plane, with all functions and their first three derivatives uniformly bounded and with (13) *satisfied over the entire plane.*

Proof. By Lemma 2, there exist four functions, $r_1(x,y)$, $\theta_1(x,y)$, $r_2(x,y)$, and $\theta_2(x,y)$ which satisfy (10) and (11). As a matter of convenience, let

$$(14) \qquad \theta(x,y) = \theta_2(x,y) - \theta_1(x,y).$$

Now let

$$(15) \qquad x = x(\xi,\eta), \qquad y = y(\xi,\eta)$$

be the conformal transformation of the exterior of G onto the exterior of Γ, the unit circle in the (ξ,η)-plane, the transformation being of class C''' both ways, even on G^* and Γ^*. The existence of such a transformation is well known [5]. Now, let $u(x,y)$ designate, in turn, each of the functions r_1, r_2, θ_1, θ_2, θ, e, and f, and let $u(\xi,\eta)$ be the transformed function on Γ^*. Define, on Γ^*,

$$u_\xi = u_x x_\xi + u_y y_\xi, \quad u_\eta = u_x x_\eta + u_y y_\eta, \quad u_\rho(1,\phi) = u_\xi \cos\phi + u_\eta \sin\phi,$$

(ρ,ϕ) being the polar coordinates in the (ξ,η)-plane, with u_x and u_y the known values on G^*, and similar definitions being made for the higher derivatives of $u(\rho,\phi)$.

Now let $h(\rho)$ be defined as a function of class C''' which is such that

$$(16)\quad h(1) = h'(1) = h''(1) = h'''(1) = h'(2) = h''(2) = h'''(2) = 0,$$

$$(17)\qquad h(2) = 1.$$

Let u represent the functions r_1, r_2, $(2/\pi)\theta$, $(2/\pi)\theta_2$, θ_1, e, and f in turn. Then for the closed annulus, $1 \leqq \rho \leqq 2$, let $u(\rho, \phi)$ be defined by

$$u(\rho, \phi) = U(\rho, \phi)[1 - h(\rho)] + Ch(\rho),$$

where the constant C is unity for r_1, $(2/\pi)\theta$, $(2/\pi)\theta_2$, e, and f, is zero for θ_1, and is ± 1 for r_2, the sign being that of r_2 on Γ^*, and where $U(\rho, \phi)$ is the function $U(x, y)$ of Lemma 3, defined for

$$u_1(x) = u(1, \phi), \qquad u_3(x) = u_{\rho\rho}(1, \phi),$$
$$u_2(x) = u_\rho(1, \phi), \qquad u_4(x) = u_{\rho\rho\rho}(1, \phi).$$

Clearly $u(\rho, \phi)$ is of class C''' for this region, and it may easily be verified that $u(\rho, \phi)$ matches up, together with its first three derivatives, with the known values on Γ^*.

From (4), (11), and (14), it is observed that r_1, r_2, and θ possess bounds on the unit circle of the form

$$| r_1 | \geqq k_1 > 0, \quad | r_2 | \geqq k_2 > 0, \quad 0 < 2m/\pi \leqq | (2/\pi)\theta | \leqq 2M/\pi < 2.$$

These bounds cannot necessarily be preserved on the annulus, but it is possible to maintain the following set of conditions there:

$$(18)\quad | r_1 | \geqq k_1/2 > 0, | r_2 | \geqq k_2/2 > 0, 0 < m/\pi \leqq | (2/\pi)\theta | \leqq (M + \pi)/\pi < 2.$$

Suppose that ρ_1 is the smallest value of ρ at which any violation of (18) occurs. Then let $h(\rho)$ satisfy (16), be monotone, and satisfy

$$h(\rho) = 1 \text{ for } (1 + \rho_1)/2 \leqq \rho = 2.$$

Clearly this preserves the bounds of (18). The other functions, together with all derivatives, are evidently bounded on the closed annulus, since each is continuous on a closed set.

For $\rho > 2$, let $u(\rho, \phi)$ and its first three derivatives be defined to possess their respective values at $\rho = 2$. As these values are all constants, $u(\rho, \phi)$ is clearly of class C''' for $\rho > 2$, and all functions, together with their derivatives, are bounded. In particular,

$$(2/\pi)\theta(2, \phi) = r_1(2, \phi) = | r_2(2, \phi) | = 1,$$

so that (13) is seen to be satisfied for all $\rho > 2$.

Now let the inverse transformation of (15) be applied. The functions all return to their original values on G^* and all derivatives match up on G^*. Outside of G^*, the functions are of class C''' and are uniformly bounded, as are all derivatives. In particular, $(2/\pi)\theta$, r_1, and r_2 are bounded by the transformed conditions (18), and thus (13) is satisfied everywhere.

As an additional result of this lemma, due to the nature of the boundary of G, there exists a square having G^* in its interior, and having the functions θ_1 and θ_2 equal to 0 and $\pi/2$, respectively, on the boundary and outside of the square.

LEMMA 5. *Let the curve C,*

$$x = x(s), \quad y = y(s), \quad a \leq s \leq a + k,$$

s being the arc length, be a simple closed curve with no cusps, and be such that the interval $(a, a + k)$ may be divided into a finite number of subintervals, (s_{i-1}, s_i) in each of which closed intervals $x(s)$ and $y(s)$ are of class C'''. Let $x(s)$ and $y(s)$ be defined for all values of s by:

$$x(s \pm k) = x(s), \quad y(s \pm k) = y(s).$$

Then there exists a function, $\theta(s)$, uniquely determined except possibly for a multiple of 2π, which is such that

$$\begin{gathered} |\,\theta([a+k]^+) - \theta(a^+)\,| = 2\pi, \\ \cos\theta(s) = x'(s), \quad \sin\theta(s) = y'(s), \quad s \neq s_i, \\ \cos\theta(s_i^+) = x'(s_i^+), \quad \sin\theta(s_i^+) = y'(s_i^+), \\ \cos\theta(s_i^-) = x'(s_i^-), \quad \sin\theta(s_i^-) = y'(s_i^-), \\ |\,\theta(s_i^+) - \theta(s_i^-)\,| < \pi. \end{gathered}$$

Proof. This is a well known theorem, and may be found in [7].

LEMMA 6. *Let $\theta(x, y)$ be of class C''' over the entire plane with $|\theta|$, $|\theta_x|$, and $|\theta_y|$ all bounded by M for all (x, y). Let $x = x(s)$, $y = y(s)$, s being the arc length, be the solution of*

$$dx/ds = \cos\theta, \qquad dy/ds = \sin\theta, \tag{19}$$

for which

$$x(0) = x_0, \qquad y(0) = y_0.^2$$

Then

---

[2] The existence and uniqueness of these functions is guaranteed by well known existence theorems in the theory of ordinary differential equations, and may be found in [1].

$$x(s_1) = x(s_2), \qquad y(s_1) = y(s_2) \tag{20}$$

implies that

$$s_1 = s_2. \tag{21}$$

Proof. If one assumes (20) to be satisfied and (21) to be not satisfied, a contradiction can be readily obtained through the use of Lemma 5.

LEMMA 7. *Let $\theta_1(x, y)$ and $\theta_2(x, y)$ be of class C' over the entire plane, be uniformly bounded together with their first derivatives there, and be such that*

$$0 < m \leq |\theta_2(x, y) - \theta_1(x, y)| \leq M < \pi$$

for all (x, y). Then there exist unique functions of class C''

$$x_1(s),\ y_1(s),\ x_2(t),\ \text{and}\ y_2(t),$$

such that

$$dx_1/ds = \cos\theta_1, \qquad dy_1/ds = \sin\theta_1 \tag{22A}$$

$$dx_2/dt = \cos\theta_2, \qquad dy_2/dt = \sin\theta_2, \tag{22B}$$

and such that

$$x_1(0) = x_1,\ y_1(0) = y_1,\ x_2(0) = x_1,\ y_2(0) = y_1.$$

For these functions,

$$x_1(s_1) = x_2(t_2),\ \text{and}\ y_1(s_1) = y_2(t_2)$$

imply that

$$s_1 = t_2 = 0.$$

Proof. This lemma follows as a consequence of the two preceding lemmas.

LEMMA 8. *Let $\theta_1(x, y)$ and $\theta_2(x, y)$ be of class C''' over the entire plane with $|\theta_1|$, $|\theta_2|$, $|\theta_{1x}|$, $|\theta_{1y}|$, $|\theta_{2x}|$, and $|\theta_{2y}|$ all bounded by $\bar{M}$ for all (x, y). Suppose also that*

$$0 < m \leq |\theta_2(x, y) - \theta_1(x, y)| \leq M < \pi$$

for all (x, y). Let $x_1(s)$ and $y_1(s)$ satisfy (22A) of Lemma 7, and let $x_1(s, t)$ and $y_1(s, t)$ be defined for each s by

$$(\partial/\partial t)[x_1(s, t)] = \cos\theta_2, \qquad (\partial/\partial t)[y_1(s, t)] = \sin\theta_2$$

with

$$x_1(s, 0) = x_1(s), \qquad y_1(s, 0) = y_1(s).$$

Then $x_1(s, t)$ and $y_1(s, t)$ are of class C''' and the transformation

$$x = x_1(s, t), \qquad y = y_1(s, t)$$

is one-to-one from the entire (s, t)-plane into a subset of the (x, y)-plane.

Proof. This follows readily from Lemma 7.

LEMMA 9. *Let $\theta(x, y)$ be of class C''' over the entire plane with $|\theta|$, $|\theta_x|$, and $|\theta_y|$ all bounded by M for all (x, y). Let*

$$x = x(s), \qquad y = y(s),$$

s being the arc length, be the solutions of

$$dx/ds = \cos\theta, \qquad dy/ds = \sin\theta,$$

for which

$$x(0) = x_0, \qquad y(0) = y_0.$$

Then for each R there exists an $s_1 > 0$ and an $s_2 > 0$ such that

$$[x(s_1) - x_0]^2 + [y(s_1) - y_0]^2 \geq R^2, \tag{23}$$

$$[x(-s_2) - x_0]^2 + [y(-s_2) - y_0]^2 \geq R^2. \tag{24}$$

Proof. Let $x_1(s, t)$ and $y_1(s, t)$ be determined by

$$\partial x/\partial t = -\sin\theta, \qquad \partial y/\partial t = \cos\theta$$

with

$$x_1(s, 0) = x(s), \qquad y_1(s, 0) = y(s). \tag{25}$$

Then by Lemma 8, the transformation

$$x = x_1(s, t), \qquad y = y_1(s, t)$$

is one-to-one from the entire (s, t)-plane into a subset of the (x, y)-plane. Let

$$J_1(s, t) = x_{1s}y_{1t} - x_{1t}y_{1s},$$

so that

$$J_1(s, 0) = 1.$$

Also, since

$$J_{1t} = -(\theta_x \cos\theta + \theta_y \sin\theta) \cdot J_1,$$

it follows that

$$e^{-2M|t|} \leq J_1(s, t) \leq e^{2M|t|}. \tag{26}$$

In addition,

$$\{x_{1t}^2 + y_{1t}^2\}^{\frac{1}{2}} = 1.$$

Now, let

$$A(s) = \int_{-T}^{+T} \int_0^s J_1(s, t)\,ds\,dt.$$

Integrating this and making use of (26),

$$A(s) \geq (s/M)(1 - e^{-2TM}),$$

and so

$$\lim_{s\to\infty} A(s) = +\infty \tag{27}$$

In particular, if $T = 1$,

$$A(s) \geq (s/M)(1 - e^{-2M}).$$

Since (27) holds, there exists an s_1 such that the area in the (x, y)-plane corresponding to the set

$$0 \leq s \leq s_1, \qquad -1 \leq t \leq 1$$

exceeds $(R+1)^2$. That is, there exists a point (s_0, t_0) in this set such that

$$[x_1(s_0, t_0) - x_0]^2 + [y_1(s_0, t_0) - y_0]^2 > (R+1)^2.$$

In addition,

$$[x_1(s_0, t_0) - x_1(s_0, 0)]^2 + [y_1(s_0, t_0) - y_1(s_0, 0)]^2 \leq 1$$

so that

$$[x_1(s_0, 0) - x_0]^2 + [y_1(s_0, 0) - y_0]^2 > R^2.$$

Thus, in view of (25), it is evident that (23) holds. A similar method of attack will demonstrate the validity of the inequality (24).

LEMMA 10. *Let G be a region bounded by a simple, closed regular curve of class C_a''', and let $a(x,y)$, $b(x,y)$, $c(x,y)$, and $d(x,y)$ be functions of class C''' on $\bar{G}$, with*

$$[a(x,y)][d(x,y)] - [b(x,y)][c(x,y)] \neq 0$$

at each point of $\bar{G}$. Then there exists a one-to-one transformation

$$\xi = \xi(x,y), \qquad \eta = \eta(x,y) \tag{28}$$

of the entire (x,y)-plane into the entire (ξ,η)-plane, which, together with its inverse, is of class C''' over the entire plane, and in which ξ and η satisfy

$$\begin{aligned} \xi_x\eta_y - \xi_y\eta_x &\neq 0, \\ a\eta_x + c\eta_y &= 0, \\ b\xi_x + d\xi_y &= 0. \end{aligned} \tag{29}$$

at each point of $\bar{G}$.

Proof. Let the extension of Lemma 4 be applied to these four given functions. Then the functions $r_1(x,y)$, $\theta_1(x,y)$, $r_2(x,y)$ and $\theta_2(x,y)$, which are the functions of Lemma 2, are all of class C''' over the entire plane, and, together with their first three derivatives, are uniformly bounded everywhere.

In particular, there exists a square $S: [(x_0, y_0), (x_0 + k, y_0 + k)]$ which contains $\bar{G}$ in its interior, and which is such that $\theta_1(x, y)$ and $\theta_2(x, y)$ equal 0 and $\pi/2$, respectively, on S^* and at all points exterior to S^*, as well as at all points (x, y) interior to S which are such that their distance from S^* is not greater than unity.

Let

$$x = x_1(s), \; y = y_1(s), \text{ and } x = x_2(t), \; y = y_2(t)$$

be the families of curves satisfying (22A) and (22B), respectively, and designate by C_1 and C_2 the curves of the first and the second families, respectively, through (x_0, y_0). Now let P: (x, y) be any point, and let C'_1 and C'_2 be the curves of the first and second families through P. By means of Lemma 9, we can easily show that C'_1 intersects C_2, and that C'_2 intersects C_1. Let P_1 and P_2 be, respectively, these two points of intersection. Then let s be the directed distance along C_1 from P to P_1, and t be the directed distance along C_2 from P to P_2. Let $\xi(x, y)$ and $\eta(x, y)$ be defined by

$$\xi(x, y) = s, \qquad \eta(x, y) = t.$$

By Lemma 8 and the above, the transformation

$$\xi = \xi(x, y), \qquad \eta = \eta(x, y)$$

is one-to-one from the entire (x, y)-plane into the entire (ξ, η)-plane.

Let $x_1(s, t)$ and $y_1(s, t)$ be defined by

$$(\partial/\partial s)[x_1(s, t)] = \cos\theta_1(x_1, y_1), \qquad (\partial/\partial s)[y_1(s, t)] = \sin\theta_1(x_1, y_1)$$

with

$$x_1(0, t) = x_2(t), \qquad y_1(0, t) = y_2(t),$$

$x = x_2(t)$, $y = y_2(t)$ being the equations of C_2. Let $x_2(s, t)$ and $y_2(s, t)$ be defined by

$$(\partial/\partial t)[x_2(s, t)] = \cos\theta_2(x_2, y_2), \qquad (\partial/\partial t)[y_2(s, t)] = \sin\theta_2(x_2, y_2)$$

with

$$x_2(s, 0) = x_1(s), \qquad y_2(s, 0) = y_1(s),$$

$x = x_1(s)$, $y = y_1(s)$ being the equations of C_1.

By Lemma 8, all of the functions are of class C''' in both s and t. Also, it follows that both of the transformations,

$$x = x_i(s, t), \qquad y = y_i(s, t), \qquad (i = 1, 2),$$

carry the entire (s, t)-plane into the entire (x, y)-plane.

If these transformations are reversed so as to obtain

$$(30) \qquad s = s_1(x, y), \ t = t_1(x, y), \text{ and } s = s_2(x, y), \ t = t_2(x, y),$$

from the first and second, respectively, then it is seen that

$$\xi = \xi(x, y) = s_2(x, y), \qquad \eta = \eta(x, y) = t_1(x, y).$$

Now, let

$$J_i(s, t) = x_{is}y_{it} - x_{it}y_{is}, \qquad (i = 1, 2);$$

it may easily be verified that $J_1(0, t)$ is never zero, and that since

$$J_{1s} = [(\partial/\partial y) \sin\theta_1 + (\partial/\partial x) \cos\theta_1] \cdot J_1,$$

$J_1(s, t)$ is never zero. Similarly, $J_2(s, t)$ is never zero. Thus the inverse transformations (30) are of class C''', and hence both ξ and η are of class C'''.

Now let

$$J(s, t) = \xi_x\eta_y - \xi_y\eta_x.$$

This may easily be reduced to

$$J = \sin[\theta_2(x, y) - \theta_1(x, y)]/J_1J_2.$$

Thus J is defined everywhere and never vanishes. Hence the inverse of (28) is of class C''' everywhere.

Now,

$$\begin{aligned} a\eta_x + c\eta_y &= r_1 \cos\theta_1\eta_x + r_1 \sin\theta_1\eta_y \\ &= r_1[x_{1s}\eta_x + y_{1s}\eta_y] \equiv 0, \end{aligned}$$

since J_1 never vanishes and since

$$\eta_x = -y_{1s}/J_1, \qquad \eta_y = x_{1s}/J_1.$$

Similarly, it may be shown that

$$b\xi_x + d\xi_y = 0.$$

This is the transformation sought for in this section. The essential features that will be needed later on are its differentiability properties, the three statements of (28), and the fact that it is one-to-one from the entire (x, y)-plane into the entire (ξ, η)-plane.

3. A lemma of general interest in the calculus of variations. In this section, we seek a lemma which is of quite general interest in the Calculus of Variations. It is necessary to develop several preliminary lemmas which yield, upon proper application of the transformation of the preceding section, the desired result.

LEMMA 11. *Let $B(x,y)$, $\alpha(x,y)$, and $\epsilon(x,y)$ be functions of class C'' on the closed rectangle $\bar{R}$: $[(x_1,y_1),(x_2,y_2)]$, with $B(x,y)$ and $\alpha(x,y)$ never zero on $\bar{R}$. Then there exists a function $u(x,y)$ which is of class C''' in x and of class C'' in y which possesses the boundary values*

$$u(x,y_1)=u(x,y_2)=u_y(x,y_1)=u_y(x,y_2)=0,$$

and which is such that

$$\int_{y_1}^{y_2} B(x,y)[\alpha(x,y)u_x(x,y)+\epsilon(x,y)u(x,y)]dy \overset{x}{\equiv} 1.$$

Proof. Let $v(x,y)$ be of class C' and be such that

$$\alpha(x,y)v_x(x,y)+\epsilon(x,y)v(x,y)=1/B(x,y).$$

If now, $\phi(y)$ is any function of class C'' satisfying

$$\phi(y_1)=\phi(y_2)=\phi'(y_1)=\phi'(y_2)=0$$

and such that

$$\int_{y_1}^{y_2}\phi(t)\,dt\neq 0,$$

then obviously the required function of the lemma is

$$u(x,y)=v(x,y)\phi(y)/\int_{y_1}^{y_2}\phi(t)\,dt.$$

LEMMA 12. *Let $\phi(x,y)$ be a function of class C' on $\bar{R}$, $\bar{R}$ being the rectangle: $[(x_1,y_1),(x_2,y_2)]$, and suppose that*

$$\int_R\int \phi(x,y)u(x,y)\,dydx=0 \tag{31}$$

for all $u(x,y)$ which are of class C'', with $u_{xxx}(x,y)$ also continuous and zero on R^*, which vanish together with their first and second derivatives on R^*, and which satisfy

$$\int_{y_1}^{y_2} B(x,y)[\alpha(x,y)u_x(x,y)+\epsilon(x,y)u(x,y)]dy \overset{x}{\equiv} 0, \tag{32}$$

$B(x,y)$, $\alpha(x,y)$, and $\epsilon(x,y)$ satisfying the conditions stated for them in the preceding lemma.

Then there exists a function $\chi(x)$ of class C' for $x_1\leq x\leq x_2$ such that

$$\phi(x,y)+(\partial/\partial x)[B(x,y)\alpha(x,y)\chi(x)]-B(x,y)\,\epsilon\,(x,y)\chi(x)=0 \tag{33}$$

on $\bar{R}$.

Proof. Let $u_1(x,y)$ be the function of Lemma 11. Now let $v(x,y)$ be any function of class C''' which vanishes together with its derivatives on R^*, and define $\psi(x)$ by

$$\psi(x) = \int_{y_1}^{y_2} B(x,t)[\alpha(x,t)v_x(x,t) + \epsilon(x,t)v(x,t)]dt.$$

Thus $\psi(x)$ is of class C'', and satisfies

$$\psi(x_1) = \psi(x_2) = \psi'(x_1) = \psi'(x_2) = 0.$$

Now let

$$v(x,y) = u^*(x,y) + u_1(x,y)\psi(x).$$

Obviously $u^*(x,y)$ satisfies both (31) and (32). If, now,

$$\int_{x_1}^{x_2}\int_{y_1}^{y_2} v(x,y)\phi(x,y)\,dydx$$

be manipulated by various interchanges of letters and orders of integration, as well as by integration by parts, there results the equation,

$$\text{(34)}\qquad \int_{x_1}^{x_2}\int_{y_1}^{y_2} v(x,y)\{\phi(x,y) - B(x,y)\,\epsilon\,(x,y)\chi(x) + (\partial/\partial x)[B(x,y)\alpha(x,y)\chi(x)]\}dydx = 0.$$

where

$$\chi(x) = \int_{y_1}^{y_2} \phi(x,t)u_1(x,t)\,dt.$$

Since (34) is satisfied by all $v(x,y)$ of class C'' which vanish together with their first and second derivatives on R^*, then the statement (33) follows immediately from the fundamental lemma of the calculus of variations.

LEMMA 13. *Let $A(x,y)$, $B(x,y)$, $C(x,y)$, $D(x,y)$, $E(x,y)$, and $F(x,y)$ be functions of class C'' on a simply connected region Γ with $AD - BC \neq 0$ on Γ. Then a necessary and sufficient condition that there exists a solution $\mu \neq 0$ of class C' on Γ of the two equations*

$$\text{(35)}\quad (\partial/\partial x)(A\mu) + (\partial/\partial y)(C\mu) = E\mu, \quad \textit{and} \quad (\partial/\partial x)(B\mu) + (\partial/\partial y)(D\mu) = F\mu$$

is that

$$\text{(36)}\qquad \frac{\partial}{\partial y}\left[\frac{D(E - A_x - C_y) - C(F - B_x - D_y)}{AD - BC}\right] = \frac{\partial}{\partial x}\left[\frac{A(F - B_x - D_y) - B(E - A_x - C_y)}{AD - BC}\right]$$

on Γ.

Proof. If the equations (35) be solved simultaneously for μ_x and μ_y, it is found that

$$\mu_x = \frac{D(E - A_x - C_y) - C(F - B_x - D_y)}{AD - BC} \cdot \mu,$$

$$\mu_y = \frac{A(F - B_x - D_y) - B(E - A_x - C_y)}{AD - BC} \cdot \mu,$$

From these, we may conclude that if $\mu \neq 0$ on Γ, then μ is never zero on Γ, and so $\log \mu$ is defined and is of class C'' on Γ. Dividing these last two equations by μ, it is clear that, if (35) holds, then (36) is also satisfied.

Finally, if (36) holds, there exists a function $\phi(x, y)$ of class C'' on Γ such that

$$\phi_x = \frac{D(E - A_x - C_y) - C(F - B_x - D_y)}{AD - BC},$$

$$\phi_y = \frac{A(F - B_x - D_y) - B(E - A_x - C_y)}{AD - BC},$$

since Γ is simply connected. Thus any function μ of the form

$$\mu = Ce^{\phi},$$

C being any constant, satisfies (35).

LEMMA 14. *Let* $a(x, y)$, $b(x, y)$, $c(x, y)$, $d(x, y)$, $e(x, y)$, $f(x, y)$, $\alpha(x, y)$, $\delta(x, y)$, $\epsilon(x, y)$, *and* $\zeta(x, y)$ *be of class* C'' *on the rectangle* $R\colon [(x_1, y_1), (x_2, y_2)]$ *with* $\alpha \cdot \delta \neq 0$ *on* $\bar{R}$. *Let* $u(x, y)$ *and* $v(x, y)$ *be of class* C' *on* $\bar{R}$, *vanish on* R^*, *and satisfy*

$$\alpha u_x + \delta v_y + \epsilon u + \zeta v = 0 \tag{37}$$

on $\bar{R}$.

Then

$$\int_R\int (au_x + bv_x + cu_y + dv_y + eu + fv)\,dydx = \int_R\int u[(\partial/\partial x)(\alpha H) - \epsilon H - (a_x + c_y - e)]\,dydx, \tag{38}$$

where

$$H(x, y) = -\frac{1}{\delta(x, y)} \exp\left[\int_{y_1}^{y} \frac{\zeta(x, t)}{\delta(x, t)}\,dt\right] \times \int_{y}^{y_2} \exp\left[-\int_{y_1}^{t} \frac{\zeta(x, r)}{\delta(x, r)}\,dr\right] \cdot [b_x(x, t) + d_y(x, t) - f(x, t)]\,dt.$$

Proof. If $u(x, y)$ and $v(x, y)$ satisfy (37) on $\bar{R}$ and vanish on R^*, we have

$$v(x, y) = K(x, y) \int_{y_1}^{y} B(x, t)[\alpha(x, t)u_x(x, t) + \epsilon(x, t)u(x, t)]\,dt, \tag{39}$$

where

$$K(x,y) = -\exp\left[-\int_{y_1}^{y} \frac{\zeta(x,r)}{\delta(x,r)}\,dr\right],$$

and

$$B(x,y) = \frac{1}{\delta(x,y)} \exp\left[\int_{y_1}^{y} \frac{\zeta(x,r)}{\delta(x,r)}\,dr\right].$$

Now if the left hand member of (38) be integrated by parts, and $v(x,y)$ be replaced by its equal from (39), there will result, after various interchanges of the order of integration as well as several integrations by parts, the right hand side of (38), as required.

LEMMA 15. *Let $a(x,y)$, $b(x,y)$, $c(x,y)$, $d(x,y)$, $e(x,y)$, $f(x,y)$, $\alpha(x,y)$, $\delta(x,y)$, $\epsilon(x,y)$, and $\zeta(x,y)$ be of class C'' with $\alpha\cdot\delta \neq 0$ on the closed rectangle $\bar{R}$: $[(x_1,y_1),(x_2,y_2)]$. Suppose that*

$$\int_R\int (au_x + bv_x + cu_y + dv_y + eu + fv)\,dy\,dx = 0$$

for all pairs of functions $u(x,y)$ and $v(x,y)$ which are of class C'' on $\bar{R}$, which vanish with their first and second derivatives on R^*, and which satisfy

$$\alpha u_x + \delta v_y + \epsilon u + \zeta v = 0 \tag{40}$$

on $\bar{R}$.

Then there exists at least one function $\lambda(x,y)$ of class C' on $\bar{R}$ which satisfies the two equations

$$(\partial/\partial x)(a - \lambda\alpha) + (\partial/\partial y)(c) = e - \lambda\epsilon;$$
$$(\partial/\partial x)(b) + (\partial/\partial y)(d - \lambda\delta) = f - \lambda\zeta$$

on $\bar{R}$.

Proof. Let $u(x,y)$ and $v(x,y)$ be of class C'' on $\bar{R}$, vanish with their first and second derivatives on R^*, and satisfy (40). Suppose that $u_{xxx}(x,y)$ is also continuous and vanishes on R^*. Then the equation (39) of Lemma 14 holds. In order for $v(x,y)$ to vanish on R^*, it must be true that

$$\int_{y_1}^{y_2} B(x,t)\,[\alpha(x,t)u_x(x,t) + \epsilon(x,t)u(x,t)]\,dt \overset{x}{\equiv} 0. \tag{41}$$

Thus, since all of the conditions of Lemma 14 are satisfied, the result stated as equation (38) must hold for all $u(x,y)$ satisfying all the stated conditions, including (41). From Lemma 12, then, it follows that there exists a function $\chi(x)$ of class C', such that

$$(\partial/\partial x)(\alpha H) - \epsilon H - (a_x + c_y - e) + (\partial/\partial x)(\alpha B\chi) - \epsilon B\chi = 0,$$

where H and B are defined as in Lemma 14.

Now, if we define

$$\lambda(x, y) = H(x, y) + B(x, y)\chi(x),$$

it may easily be verified that this function satisfies the required conditions of the lemma.

Lemma 16. *Let G be a region bounded by a simple, closed, regular curve of class C_a'''. Let $A(x, y)$, $B(x, y)$, $C(x, y)$, $D(x, y)$, $E(x, y)$, $F(x, y)$, $a(x, y)$, $b(x, y)$, $c(x, y)$, $d(x, y)$, $e(x, y)$, and $f(x, y)$ be functions of class C''' on $\bar{G}$, with*

$$[A(x, y)][D(x, y)] - [B(x, y)][C(x, y)] \neq 0$$

on $\bar{G}$. Suppose also that the set of points (x, y) where

$$\frac{\partial}{\partial y}\left[\frac{D(E - A_x - C_y) - C(F - B_x - D_y)}{AD - BC}\right] = \frac{\partial}{\partial x}\left[\frac{A(F - B_x - D_y) - B(E - A_x - C_y)}{AD - BC}\right] \tag{42}$$

is nowhere dense. Suppose finally that

$$\iint_R (au_x + bv_x + cu_y + dv_y + eu + fv)\,dy\,dx = 0 \tag{43}$$

for all $u(x, y)$ and $v(x, y)$ which are of class C'' on $\bar{G}$, which vanish together with their first and second derivatives on G^*, and which satisfy

$$Au_x + Bv_x + Cu_y + Dv_y + Eu + Fv = 0. \tag{44}$$

Then there exists a unique function $\lambda(x, y)$, which is of class C' on $\bar{G}$, and which is such that

$$\begin{aligned} (\partial/\partial x)(a - \lambda A) + (\partial/\partial y)(c - \lambda C) &= e - \lambda E, \\ (\partial/\partial x)(b - \lambda B) + (\partial/\partial y)(d - \lambda D) &= f - \lambda F. \end{aligned} \tag{45}$$

Proof. Let the transformation of Lemma 10 be applied to these given functions. Then (43) becomes, Γ being the transform of G,

$$\iint_\Gamma (a_1 u_\xi + b_1 v_\xi + c_1 u_\eta + d_1 v_\eta + e_1 u + f_1 v)\,d\eta\,d\xi = 0 \tag{46}$$

where

$$a_1(\xi, \eta) = \{a[x(\xi, \eta), y(\xi, \eta)]\xi_x + c[x(\xi, \eta), y(\xi, \eta)]\xi_y\}(x_\xi y_\eta - x_\eta y_\xi),$$

with b_1, c_1, d_1, e_1, and f_1 represented by similar expressions, and where $u(\xi, \eta)$ and $v(\xi, \eta)$ are the transforms of $u(x, y)$ and $v(x, y)$. Equation (44) becomes

$$\begin{aligned} (A_1\xi_x + C_1\xi_y)u_\xi + (B_1\xi_x + D_1\xi_y)v_\xi + (A_1\eta_x + C_1\eta_y)u_\eta \\ + (B_1\eta_x + D_1\eta_y)v_\eta + E_1 u + F_1 v = 0, \end{aligned} \tag{47}$$

A_1, B_1, C_1, D_1, E_1, and F_1 being the transformed functions A, B, C, D, E, and F respectively. From the properties of the transformation, the coefficients of u_η and v_ξ are zero, and, in addition,

$$(A_1\xi_x + C_1\xi_y)(B_1\eta_x + D_1\eta_y) = (A_1D_1 - B_1C_1)(\xi_x\eta_y - \xi_y\eta_x) \neq 0,$$

so that (47) may be written as

$$\alpha u_\xi + \delta v_\eta + \epsilon u + \zeta v = 0,$$

α, δ, ϵ, and ζ having the obvious interpretations, with

$$\alpha(\xi,\eta)\delta(\xi,\eta) \neq 0.$$

Also $u(\xi,\eta)$ and $v(\xi,\eta)$ are of class C'' and vanish with their first two derivatives on Γ^*.

Now let $\bar{p}$ be a rectangle in $\bar{\Gamma}$, and let $\bar{R}$ be the corresponding region in $\bar{G}$. Then, by Lemma 15, there exists a function $\lambda_1(\xi,\eta)$ which is of class C' on $\bar{p}$ and which satisfies

$$\begin{aligned}(\partial/\partial\xi)(a_1 - \lambda_1\alpha) + (\partial/\partial\eta)(c_1) &= e_1 - \lambda_1\epsilon,\\ (\partial/\partial\xi)(b_1) + (\partial/\partial\eta)(d_1 - \lambda_1\delta) &= f_1 - \lambda_1\zeta.\end{aligned} \tag{48}$$

Now let a, A, etc., be the original functions of x and y with $x(\xi,\eta)$ and $y(\xi,\eta)$ substituted in when the context demands. Let

$$J = \xi_x\eta_y - \xi_y\eta_x, \qquad J_1 = x_\xi y_\eta - x_\eta y_\xi.$$

On rewriting the first of equations (48), using the formulas for a_1, α, etc., we have

$$\frac{\partial}{\partial\xi}\left[\frac{a\xi_x + c\xi_y}{J} - \lambda_1(A\xi_x + C\xi_y)\right] + \frac{\partial}{\partial\eta}\left[\frac{a\eta_x + c\eta_y}{J}\right] = e/J - \lambda_1 E.$$

Now by manipulation of this last expression, including the use of formulas for the derivatives of the inverse functions, as well as certain properties of the transformation, we obtain

$$a_x + c_y - \lambda(A_x + C_y) - A\lambda_x - C\lambda_y = e - \lambda E,$$

where

$$\lambda = \lambda_1/J_1.$$

This is the first of the equations of (45), and the second may be obtained analogously.

Thus on each region $\bar{R}$ of the type indicated, there exists at least one $\lambda(x,y)$ satisfying (45). Since the set where (42) holds is nowhere dense in $\bar{G}$, it follows that (42) cannot hold identically on $\bar{R}$, and it follows from Lemma 13 that $\lambda(x,y)$ is uniquely determined on $\bar{R}$ and hence over the interior of G.

Finally, if $\lambda_1(\xi,\eta)=\lambda(x,y)/J$, it can easily be shown that λ_1, $\lambda_{1\xi}$, and $\lambda_{1\eta}$ have continuous limits on Γ^* by studying the behavior of these functions along a small part of Γ^*.

4. The Lagrange multiplier rule. In this final section, we wish to prove the Lagrange Multiplier Rule for the case of two dependent variables and two independent variables.

Suppose, then, that $f(x,y,z_1,z_2,p_1,p_2,q_1,q_2)$ and $\phi(x,y,z_1,z_2,p_1,p_2,q_1,q_2)$ are of class C'''' in their arguments, that $\bar{z}_1(x,y)$ and $\bar{z}_2(x,y)$ are of class C'''' on a region $\bar{G}$, the boundary of which is a simple, closed, regular curve of class C_a''', and that $\bar{z}_1(x,y)$ and $\bar{z}_2(x,y)$ minimize

$$\int_G\int f(x,y,z_1,z_2,p_1,p_2,q_1,q_2)\,dydx \tag{49}$$

among all functions z_1, z_2 of class C'' on $\bar{G}$, coinciding with $\bar{z}_1$ and $\bar{z}_2$, respectively, on G^*, and satisfying

$$\phi(x,y,z_1,z_2,p_1,p_2,q_1,q_2)=0. \tag{50}$$

Now suppose for the moment that $Z_1(x,y;\mu)$ and $Z_2(x,y;\mu)$ are of class C'' in their arguments for (x,y) on $\bar{G}$ and $|\mu|<\mu_0$, coincide with $\bar{z}_1$ and $\bar{z}_2$ if $\mu=0$, and satisfy (50) for $|\mu|<\mu_0$. Then, if we differentiate with respect to μ and then set $\mu=0$, we see that

$$\bar{\phi}_{p_1}\zeta_{1x}+\bar{\phi}_{p_2}\zeta_{2x}+\bar{\phi}_{q_1}\zeta_{1y}+\bar{\phi}_{q_2}\zeta_{2y}+\bar{\phi}_{z_1}\zeta_1+\bar{\phi}_{z_2}\zeta_2=0 \tag{51}$$

and ζ_1 and ζ_2 vanish on G^*. Here we have set

$$\zeta_1=[(\partial/\partial\mu)Z_1(x,y;\mu)]_{\mu=0},\qquad \zeta_2=[(\partial/\partial\mu)Z_2(x,y;\mu)]_{\mu=0}$$

and $\bar{\phi}_{p_1}$, etc., stand for $\phi_{p_1}(x,y,\bar{z}_1,\bar{z}_2,\bar{z}_{1x},\bar{z}_{2x},\bar{z}_{1y},\bar{z}_{2y})$, etc. Equation (51) is known as the equation of variation.

We cannot prove the multiplier rule for all possible cases, but shall prove it only for the case where the surface $z_1=\bar{z}_1(x,y)$, $z_2=\bar{z}_2(x,y)$ is 'quasi-normal' with respect to the differential equation (50).

Definition. The surface $z_1=\bar{z}_1(x,y)$, $z_2=\bar{z}_2(x,y)$ is said to be *quasi-normal* with respect to (50) if

(i) $AD-BC\neq 0$ on $\bar{G}$, $(A=\bar{\phi}_{p_1},\ B=\bar{\phi}_{p_2},\ C=\bar{\phi}_{q_1},\ D=\bar{\phi}_{q_2},\ E=\bar{\phi}_{z_1},\ F=\bar{\phi}_{z_2})$;

(ii) The set where

$$\frac{\partial}{\partial y}\left[\frac{D(E-A_x-C_y)-C(F-B_x-D_y)}{AD-BC}\right] = \frac{\partial}{\partial x}\left[\frac{A(F-B_x-D_y)-B(E-A_x-C_y)}{AD-BC}\right]$$

is nowhere dense on $\bar{G}$; and

(iii) If ζ_1 and ζ_2 are any functions of class C'' on $\bar{G}$ which satisfy (51) on $\bar{G}$ and vanish on G^*, there exists a one-parameter family $Z_1(x,y;\mu)$, $Z_2(x,y;\mu)$ of pairs of functions of class C'' in their arguments for (x,y) a point of $\bar{G}$ and $|\mu| < \mu_0$ such that

(a) $Z_1(x,y;0) = \bar{z}_1(x,y)$, $Z_2(x,y;0) = \bar{z}_2(x,y)$;

(b) $[\partial Z_1/\partial\mu]_{\mu=0} = \zeta_1$, $[\partial Z_2/\partial\mu]_{\mu=0} = \zeta_2$;

(c) for each μ with $|\mu| < \mu_0$, $Z_1(x,y;\mu)$ and $Z_2(x,y;\mu)$ coincide on G^* with $z_1(x,y)$ and $z_2(x,y)$, respectively; and

(d) $Z_1(x,y;\mu)$ and $Z_2(x,y;\mu)$ satisfy (50) for each μ with $|\mu| < \mu_0$.

THEOREM 1. *Let f, ϕ, G, $\bar{z}_1$, and $\bar{z}_2$ satisfy the differentiability hypotheses of the second paragraph of this section and suppose that the surface $z_1 = \bar{z}_1(x,y)$, $z_2 = \bar{z}_2(x,y)$ is a quasi-normal surface with respect to the differential equation* (50). *Suppose that $\bar{z}_1(x,y)$ and $\bar{z}_2(x,y)$ minimize the integral* (49) *among all pairs $z_1(x,y)$, $z_2(x,y)$ of class C'' on $\bar{G}$, coinciding with $\bar{z}_1$ and $\bar{z}_2$ on G^*, and satisfying* (50).

Then there exists a unique function, $\lambda(x,y)$ of class C' on $\bar{G}$ which satisfies the equations

$$(\partial/\partial x)(\bar{f}_{p_1}-\lambda\bar{\phi}_{p_1}) + (\partial/\partial y)(\bar{f}_{q_1}-\lambda\bar{\phi}_{q_1}) = \bar{f}_{z_1}-\lambda\bar{\phi}_{z_1},$$
$$(\partial/\partial x)(\bar{f}_{p_2}-\lambda\bar{\phi}_{p_2}) + (\partial/\partial y)(\bar{f}_{q_2}-\lambda\bar{\phi}_{q_2}) = \bar{f}_{z_2}-\lambda\bar{\phi}_{z_2}$$

on $\bar{G}$. (The notation of equation (51) is employed here.)

Proof. Let $\zeta_1(x,y)$ and $\zeta_2(x,y)$ satisfy the conditions (iii) of the definition. Since the surface $z_1 = \bar{z}_1(x,y)$, $z_2 = \bar{z}_2(x,y)$ is quasi-normal with respect to (50), there exists a one-parameter family of pairs of functions $Z_1(x,y;\mu)$, $Z_2(x,y;\mu)$ which satisfy the conditions (iii) of the definition. Let $I(\mu)$ be the function of μ obtained by substitution of the $Z_i(x,y;\mu)$ in the integral (49). Since $\bar{z}_1$ and $\bar{z}_2$ minimize this integral, we must have

$$I'(0) = \int_G\int (\bar{f}_{p_1}\zeta_{1x} + \bar{f}_{p_2}\zeta_{2x} + \bar{f}_{q_1}\zeta_{1y} + \bar{f}_{q_2}\zeta_{2y} + \bar{f}_{z_1}\zeta_1 + \bar{f}_{z_2}\zeta_2)\,dydx = 0.$$

That is to say, we have

$$\int_G\int (a\zeta_{1x}+b\zeta_{2x}+c\zeta_{1y}+d\zeta_{2y}+e\zeta_1+f\zeta_2)\,dydx=0$$

where $a=f_{p_1}$, etc., for every pair (ζ_1,ζ_2) of class C'' on $\bar{G}$, vanishing on G^*, and satisfying

$$A\zeta_{1x}+B\zeta_{2x}+C\zeta_{1y}+D\zeta_{2y}+E\zeta_1+F\zeta_2=0$$

where $A=\bar{\phi}_{p_1}$, etc. The existence and uniqueness of $\lambda(x,y)$ follows from Lemma 16 of **3**, since it is clear that a, A, etc., are all of class C'''.

UNIVERSITY OF NEW MEXICO.

BIBLIOGRAPHY.

[1] L. Bieberbach, *Theorie der Differentialgleichungen*, Berlin, 1930, p. 34.

[2] G. A. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-744.

[3] Max Coral, "The Euler-Lagrange multiplier rule for double integrals," *Contributions to the Calculus of Variations*, Chicago, 1933, pp. 63-94.

[4] W. Gross, "Das isoperimetrische Problem bei Doppelintegrale," *Monatschefte für Mathematik und Physik*, vol. 27 (1916), pp. 114-120.

[5] L. Lichtenstein, "Neuere Entwicklung der Potentialtheorie, Konforme Abbildung," *Encyklopädie der Mathematischen Wissenschaften*, v. 2: 3: 1, p. 209.

[6] W. F. Osgood, *Lehrbuch der Funktionentheorie*, Leipzig, 1912, p. 140.

[7] J. Radon, "Über die Randwertaufgaben beim logarithmischen Potential," *Sitzungberichte der Akademie der Wissenschaften in Wien*, v. 128: 2A; 2, 1919, pp. 1123-1140.

THE NON-LOCAL EXISTENCE PROBLEM OF ORDINARY DIFFERENTIAL EQUATIONS.*

By AUREL WINTNER.

Introduction. Let $f_1, \cdots, f_n$ be real-valued, continuous functions of position on the $(n+1)$-dimensional region

$$(1) \qquad -a \leqq t \leqq a, \qquad -\infty < x_i < \infty; \qquad (i = 1, \cdots, n).$$

Then, corresponding to any set of real values c_i, the system of differential equations

$$(2) \qquad x'_i = f_i(t; x_1, \cdots, x_n); \qquad (i = 1, \cdots, n),$$

where $x' = dx/dt$, has *near* $t = 0$ a solution $x_i = (t)$ satisfying

$$(3) \qquad x_i(0) = c_i.$$

The italicized proviso is essential, even if the system is "conservative" (in the sense that (1) and (2) become

$$(4) \qquad -\infty < x_i < \infty; \qquad (i = 1, \cdots, n)$$

and

$$(5) \qquad x'_i = f_i(x_1, \cdots, x_n)$$

respectively) and even if the functions f_i are regular-analytic on the whole x-space.

For instance, if (5) is the single differential equation

$$(6) \qquad x' = x^2,$$

then, although $f_1 = f_n$ is a polynomial, every solution $x(t)$ determined by a non-vanishing $x(0)$ ceases to exist at a finite $t = t_0$. In fact, all solutions distinct from the solution $x(t) \equiv 0$ (which belongs to $x(0) = 0$) are represented by $x(t) = (t_0 - t)^{-1}$, where t_0 is a non-vanishing integration constant determined by the initial value $x(0)$. But the function $(t_0 - t)^{-1}$ cannot be continued beyond $t = t_0$ so as to remain a solution. For, if it could, then,

---

* Received March 28, 1945.

since it should satisfy the differential equation, it ought to be defined at $t = t_0$ in such a way as to be differentiable, hence continuous, at $t = t_0$. But this is impossible, since the solution is not bounded near $t = t_0$.

These trivial comments are made only in view of (iii bis) below and in order to make it clear that the possibility of "moving *singularities*" in the particular case of *analytic* differential equations is just a manifestation of the *local* nature of the existence theorem in the *general* case. In this regard, not even Riccati's type, though reducible to a linear differential equation, is exceptional, since (6) is of his type.

On the other hand, the existence theorem of the initial problem (3) is not of a local nature if the system (2) is linear and, without additional loss of generality, homogeneous, that is, if

$$f_i(t; x_1, \cdots, x_n) = f_{i1}(t)x_1 + \cdots + f_{in}(t)x_n, \tag{7}$$

where, by virtue of the assumption made with regard to the range (1), the n^2 functions $f_{ik}(t)$ are continuous on the interval $-a \leqq t \leqq a$. In fact, the solution of (2) determined by arbitrary initial data (3) is known to exist on the whole interval $-a \leqq t \leqq a$ in the case (7). As was shown in [1], pp. 553-556, the non-local existence theorem for the linear case can be obtained as a *corollary* of the local existence theorem for the non-linear case, if the limited t-range supplied by the latter is applied repeatedly. The possibility of this reduction depended on the divergence of the harmonic series, $\Sigma 1/m = \infty$, a series resulting from the particular structure of (7).

Loc. cit., this elimination of the non-local existence theorem of linear systems as an *independent* theorem was just a matter of convenience, namely, a way of replacing two proofs by one. It took me twenty years to observe that the reduction process applied in [1] has more to it than this, just *methodical*, interest. In fact, it turns out that the whole of this reduction process can be worded so as to avoid the assumption of linearity. And it happens that the resulting rewording leads to *a solution of the non-local existence problem ("in the large," i. e., in the "unrestricted" sense of* [3], pp. 85-86, 132, 341) *for the non-linear equations* (2); *a solution which proves to be of a final nature.* Cf. (iii*) and (iii bis) below.

Both the result and its proof are very similar to the procedure applied by me to Strömgren's empirical principle of natural termination. On the other hand, an application of Birkhoff's subsequent variant of this proof of the termination principle would not lead to a solution of the problem (that is, to (iii*), (iii), or, for that matter, to the weaker results (i*), (i), which are not final in nature), since Birkhoff's variant consists in the replacement of the *constructive* process of prolongation by the *indirect* device of the Heine-Borel theorem. For references cf. [3], pp. 441-442.

It should finally be mentioned that the proofs, which will be worded for the case of a real variable t, need no modification if t is replaced by $z = x + iy$; so that *all results remain valid in the case of (ordinary, non-linear) analytic differential equations in the complex domain.*

1. The arrangement of the proofs becomes particularly convenient if the most general case, instead of being formulated directly, is reached through successive stages of straightforward generalizations. In fact, all the essential elements of the proof are needed for the following particular case:

(i) *If* $f_1, \cdots, f_n$ *are continuous functions of position on the space* (4), *and if* $(c_1, \cdots, c_n)$ *is any point of this space, then* (5) *and* (3) *have a solution* $x_i = x_i(t)$ *which exists on the whole t-axis,* $-\infty < t < \infty$, *whenever the given functions* f_i *are subject to the n inequalities*

$$f_i(x_1, \cdots, x_n) = O(|x|) \text{ as } |x| \to \infty. \tag{8}$$

Here, and in the sequel, $|x|$ denotes, not the Euclidean length of the vector x, but the length of the longest component of x, that is,

$$|x| = \max(|x_1|, \cdots, |x_n|), \text{ where } x = (x_1, \cdots, x_n) \tag{9}$$

(this leads to $|x| = |x|$ if $n = 1$). Accordingly, (8) means the existence of a constant A satisfying

$$|f_i(x_1, \cdots, x_n)| < A \max(|x_1|, \cdots, |x_n|, 1). \tag{10}$$

The $(n+1)$-th term, 1, on the right of (10) is adjoined in order to take care of the vicinity of the origin of the x-space.

Since only the continuity of the functions f_i is assumed in (i), the solutions to which the assertion of (i) refers will not in general be unique. In addition, the solution can be unique near $t = 0$ but then lead to branch-points. What will happen in such a case is clear from the wording of (i) and from the following proof. Needless to say, the emphasis in (i) is not on this generality, since what really is interesting in (i) is the case of "smooth" functions f_i of the type occurring in dynamical applications (or, for that matter, the case of regular-analytic functions); cf. [3], p. 356. And the solutions must be unique *throughout* (and so, by (i), for $-\infty < t < \infty$), if the given functions f_i satisfy, for instance, the local condition of Lipschitz (e. g., if every f_i is of class C').

If the n continuous functions $f_i(x_1,\cdots,x_n)$ have an absolute value not exceeding M in the cube $|x_i - x_{i0}| \leqq b$ about a point $(x_{10},\cdots,x_{n0})$, and if t_0 is arbitrary, then (5) has a solution $x_i(t)$ which satisfies the initial condition $x_i(t_0) = x_{i0}$ and exists, at least, on the t-interval consisting of those values which differ from t_0 by not more than b/M, and the n inequalities $|x_i(t) - x_{i0}| \leqq b$ hold for every t contained in this interval. The lower bound b/M for the t-range of assured existence follows by the method of equicontinuous functions. The same b/M results when, under an additional restriction of Lipschitz's type, another method, such as the method of successive approximations, is applied. Actually, *b/M cannot be improved in the complex-analytic case*, a case in which b/M becomes a "best universal constant" (of the type of absolute constants occurring in the theory of conformal mapping); cf. [2].

If the constants x_{i0}, t_0, b are identified with $c_i, 0, 1$ respectively, it is clear from the definition of M and from (10), that M can be chosen to be $A \max(c, 1) = Ac$, if c is any value exceeding each of the $n+1$ numbers $|c_i|, 1$. Hence, if c is fixed in this manner, then $b/M = 1/M$ becomes the reciprocal value of Ac. Consequently, if

$$t_1 = A^{-1}/c, \tag{11}$$

then (2) and (3) have a solution $x_i(t)$ on the interval $0 \leqq t \leqq t_1$, and the n inequalities $|x_i(t) - c_i| \leqq 1$ hold for every t contained in this interval and so, in particular, for $t = t_1$. It follows, therefore, from the definition of c that

$$|x_i(t_1)| < c + 1. \tag{12}$$

Next, let x_{i0}, t_0, b be identified with $x_i(t_1), t_1, 1$ respectively. Then it is clear from the definition of M (for the resulting cube) and from (10) and (12), that $A \max(c+1, 1) = A(c+1)$ is an admissible value of M. Hence, $b/M = 1/M$ becomes t_2, if t_2 denotes the quotient

$$t_2 = A^{-1}/(c+1). \tag{13}$$

Consequently, the solution $x_i(t)$, the existence of which was assured for $0 \leqq t \leqq t_1$, proves to be prolongable for $t_1 \leqq t \leqq t_1 + t_2$. Since $x_{i0} = x_i(t_1)$, it also follows that the n inequalities $|x_i(t) - x_i(t_1)| \leqq 1$ are satisfied for every t contained in the latter interval and so, in particular, for $t = t_1 + t_2$. Accordingly, from (12),

$$|x_i(t_1 + t_2)| < c + 2. \tag{14}$$

If x_{i0}, t_0, b are identified with $x_i(t_1+t_2)$, t_1+t_2, 1 respectively, it becomes clear that the possibility of repeating this procedure is never arrested. The m-th term of the sequence which starts with (11) and (13) is seen to be

$$t_m = A^{-1}/(c+m-1), \tag{15}$$

and the m-th step extends the solution initiated by (3) to the t-interval

$$t_1+\cdots+t_{m-1} \leqq t \leqq t_1+\cdots+t_m, \qquad (t_0=0), \tag{16}$$

since, corresponding to (12) and (14),

$$|x_i(t_1+\cdots+t_{m-1})| < c+m-1. \tag{17}$$

The first m intervals (16) cover the range $0 \leqq t \leqq T_m$, where

$$T_m = t_1+\cdots+t_m. \tag{18}$$

But, since $\Sigma 1/m = \infty$, it is clear from (15) and (18) that $T_m \rightarrow \infty$ as $m \rightarrow \infty$. Hence, the existence domain of the solution $x_i(t)$ contains the whole half-axis $0 \leqq t < \infty$. Since t can be replaced by $-t$, the proof of (i) is now complete.

2. If the functions f_i are polynomials, then the assumption (8) of (i) is satisfied only if the system (5) is linear. On the other hand, (8) includes an extensive class of algebraic, and for that matter rational, functions f_i, as well as entire functions f_i (as exemplified by such types as

$$f(x) = (1+x^3)^{-\frac{1}{2}}, \quad (1+x^2)^{1/3}, \quad x^3/(1+x^2), \quad x \exp \sin x,$$

where $n=1$). In the linear case, the general solution of (5) consists of exponentials e^{at} and (possibly) of $t, t^2, \cdots, t^{n-1}$, since t does not occur in (5). In particular, every solution is $O(e^{Ct})$, as $t \rightarrow \infty$, in this trivial case. In the non-linear case, no "explicit" solution of (5) is possible, of course (except when $n=1$). But it turns out that *the growth of the solutions in the non-linear case* is limited by the same estimate as in the trivial case of a linear system with constant coefficients:

(ii) *Under the assumptions of* (i), *each of the n components $x_i(t)$ of every solution of* (5) *is $O(e^{Ct})$ as $t \rightarrow \infty$, where C is a constant. In addition, C can be chosen so as to be independent of the integration constants* (3).

In fact, it is clear from the proof of (17) that $|x_i(t)| < c+m$ holds for every positive t not exceeding the value (18). On the other hand, it is seen from (15) that, as $m \to \infty$, the value (18) is asymptotically equal to $A^{-1} \log m$. This proves the first part of (ii). The refinement claimed by the second part follows by observing that the constant A, being defined by the formulation (10) of (8), depends only on the system (5).

The trivial case mentioned before (ii) shows that the estimate supplied by (ii) cannot be improved if only (8) is assumed. However, it is clear from the proof of (ii) that C in (ii) can be chosen arbitrarily small if (10) holds for every A. In other words, (ii) can be supplemented as follows:

(ii bis) *If* $O(|x|)$ *in the assumption* (8) *of* (i) *is refined to* $o(|x|)$, *then* $O(e^{Ct})$ *in the assertion of* (ii) *can be improved to* $\exp o(t)$.

By an adaptation of these remarks, estimates of the growth which correspond to (ii), (ii bis) can readily be obtained for each of the cases covered by the general result (iii) below.

3. In (i), the differential equations are of the particular type (5), that is, t does not occur explicitly. In order to extend (i) to the general case (2), let $f_1, \cdots, f_n$ be given as continuous functions of position on the $(n+1)$-dimensional region (1). The t-bound b/M supplied by the local existence theorem, used in the above proof of (i), must then be replaced by $\min(a, b/M)$, where a is the number occurring in (1). Correspondingly, the values (11), (13), (15) must now be restricted so as to be within the t-range of (1), that is, (15) must be replaced by

$$t_m = \min(a, A^{-1}/(c + m - 1)). \tag{19}$$

Then the extension of the solution to the consecutive intervals (16) is exactly the same as in the proof of (i), if it is assumed that the functions f_i (which now contain t) satisfy the inequalities (10), where A is a constant (independent of $(x_1, \cdots, x_n)$ *and* of t).

Thus, under the assumption just mentioned, the existence of the solution $x_i(t)$ follows on the interval $0 \leqq t \leqq T_m$, where m is arbitrary and T_m is defined by (18) and (19). However, since $\Sigma 1/m = \infty$, it is clear from (18) and (19) that T_m becomes identical with a when m is large enough. Consequently, (i) can be extended from (5) to (2), as follows:

(i*) *If* $f_1, \cdots, f_n$ *are continuous functions of the position* $(t; x_1, \cdots, x_n)$ *on the domain* (1), *and if* $c_1, \cdots, c_n$ *are arbitrary, then* (2) *and* (3) *have a solution* $x_i = x_i(t)$ *on the whole t-interval,* $-a \leqq t \leqq a$, *whenever the given functions are subject to the n inequalities*

$$\max_{-a \leqq t \leqq a} |f_i(t; x_1, \cdots, x_n)| = O(|x|) \tag{20}$$

as (9) *tends to* ∞.

Clearly, (i*) implies the classical result as to the unrestricted existence of the solutions of linear differential equations. In fact, (20) is satisfied in the case (7), where the $f_{ik}(t)$ are arbitrary continuous functions on the interval $-a < t \leqq a$. Conversely, since the estimate (20) is just what is required in the case of linear systems, the exceptional function-theoretical standing of the linear systems might suggest that (20) cannot be relaxed. However, it turns out that (20) is not the true condition, since (i*) can be improved as follows:

(iii*) *The assertion of* (i*) *remains valid if* $O(|x|)$ *in* (20) *is relaxed to* $O(L(|x|))$, *where* $L(r)$ *is any L-function satisfying*

$$\int^{\infty} dr/L(r) = \infty. \tag{21}$$

This contains (i*), since $\int dr/L(r) = \log r$ in the case, $L(r) = r$, of (20). And (iii*) is the true theorem, since its integral condition cannot be altered, as will be shown below.

4. The proof of (i), where (2) is of the form (5), is based on the fact that the series $\Sigma 1/m$ diverges. Correspondingly, the proof of (i) remains unaltered in cases in which $\Sigma 1/(m \log m)$ or $\Sigma 1/(m \log m \log\log m)$ takes the place of $\Sigma 1/m$. On the other hand, what introduced the particular series $\Sigma 1/m$ in the proof of (i) was just the linearity of the function, (9), occurring in the O-assumption, (8), of (i). Thus it is clear that (i) can be generalized as follows:

(iii) *The assertion of* (i) *remains valid if* $O(|x|)$ *on the right of* (8) *is relaxed to* $O(L(|x|))$, *where* $L(r)$ *is any L-function satisfying* (21).

Actually, the assumption that $L(r)$ be an L-function is not needed to its full extent. On the other hand, the assumption (21) is fundamental indeed:

(iii bis) *Corresponding to any L-function subject to*

$$\int^{\infty} dr/L(r) < \infty, \tag{21 bis}$$

there exist systems (5), *even analytic systems, which satisfy the assumptions of* (iii) *except the assumption* (21), *but are such as to make the assertion of* (iii) *false* (*i. e., such as to possess solutions* $x_i(t)$ *which cease to exist at a finite* $t = t^0$).

In order to see this, let $n = 1$, and let $f_1(x_1) = f(x)$ be identical with $L(x)$ (for large positive $x_1 = x$), where $L(r)$ is any given L-function satisfying (21 bis). Thus the system (5) becomes the single equation $x' = L(x)$ (for large $x > 0$). Its solution $x = x(t)$ results by the inversion of the relation

$$t = \int_c^x dr/L(r) + \text{const.} \tag{22}$$

(if c is large enough). Since $L > 0$, the integral (22) is an increasing function of x ($> c$), and so the correspondence (22) between t and x is one-to-one. But it is clear from (22) and (21 bis) that t tends to a finite limit as $x \rightarrow \infty$. Hence, if t^0 denotes this limit, then $x(t) \rightarrow \infty$ as $t \rightarrow t^0$, where $t < t_0$ ($< \infty$). It follows therefore from the remarks made after (6), that the solution $x(t)$ cannot be prolonged beyond t^0.

This proves (iii bis). And (iii bis) implies the assertion made after (iii*). Finally, in order to prove (iii*), it suffices to make in the proof of (iii) the trivial modification which led from the proof of (i) to that of (i*).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Wintner, "Zur Theorie der unendlichen Differentialsysteme," *Mathematische Annalen*, vol. 95 (1925), pp. 544-556.

[2] ———, "On the exact value of the bound for the regularity of solutions of ordinary differential equations," *American Journal of Mathematics*, vol. 57 (1935), pp. 539-540.

[3] ———, *The Analytical Foundations of Celestial Mechanics*, Princeton, 1941.

CRITERIA FOR THE ABSOLUTE CONVERGENCE OF A FOURIER SERIES AT A GIVEN POINT.*

By KIEN-KWONG CHEN.

1. In the case of the convergence of the Fourier series of f at a given point, the problem is entirely controlled by the behavior of f in the neighborhood of that point. But the absolute convergence of the series is not a local property, and in fact depends on the behavior of the function throughout the whole interval of definition.

Various sufficient conditions for the absolute convergence of the series

$$\tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{1}$$

are known.[1] Any one of these implies the convergence of the series

$$\Sigma (| a_n | + | b_n |). \tag{2}$$

The necessary and sufficient condition for the absolute convergence of a trigonometrical series in the whole interval, is that the series is a Fourier series of a Young's continuous function.[2]

2. Let $h(x)$ be an even function such that $h(x) = x^a$ $(0 < a < 1)$ for $0 \leqq x \leqq \pi$, and that $h(x + 2\pi) = h(x)$ for every x. Write

$$g(x) = \sum_{n=1}^{\infty} (\sin nx)/n, \qquad h(x) + g(x) = f(x),$$

and suppose that the Fourier series of $f(x)$ is (1). Then we have $b_n = 1/n$. An easy calculation shows that $a_n = O(n^{-1-a})$. Thus, the Fourier series of f converges absolutely at the point

$$x = 0, \tag{3}$$

without the convergence of the series (2). It is perhaps worth remarking that the example may be obtained more quickly by setting $a_n = 0$, $b_n = 1/n$.

---

* Received February 21, 1944.

[1] Cf. A. Zygmund (1), Chapter VI.

[2] K. K. Chen (1); G. H. Hardy and J. E. Littlewood (1).

The absolute convergence at the origin (3) of the Fourier series can however be known by the following criterion.[3].

THEOREM 1. *If* $0 < p < 1$, *and the derivative of*

$$\int_0^t \frac{u^p \phi(u)\,du}{(t-u)^p} \tag{4}$$

exists as a function of bounded variation i nthe interval $0 \leqq t \leqq \pi$, *where*

$$\phi(t) = \tfrac{1}{2}\{f(x+t) + f(x-t)\} \tag{5}$$

then the Fourier series of f *converges absolutely at the point* x.

In fact, when $\phi(t) = \frac{1}{2}t^a$, the derivative of the function (4) becomes a constant multiple of t^a which is of bounded variation in $(0, \pi)$, since a is non-negative.

3. In order to prove Theorem 1, we establish several lemmas.

Suppose that $0 < \alpha < 1$, $x \geqq 0$, and that the integral of h from 0 to x, of order $1-\alpha$

$$(h(x))_{1-\alpha} = (1/\Gamma(1-\alpha)) \int_0^x (x-t)^{-\alpha} h(t)\,dt$$

has the derivative $(d/dx)(h(x))_{1-\alpha}$ at the point x, then we write

$$(h(x))_{-\alpha} = (d/dx)(h(x))_{1-\alpha}.$$

LEMMA 1. *If the derivative of* $(h(x))_{1-\alpha}$ $(0 < \alpha < 1)$ *exists as a bounded function in* $(0, X)$, *then* $(h(x))_{\alpha}$ *is equivalent to a function of* Lip α *and*

$$h(x) = ((h(x))_{-\alpha})_{\alpha}, \qquad (0 \leqq x \leqq X),$$

except for a null set of points x.

Proof. Write

$$(d/dx)(h(x))_{1-\alpha} = g(x), \qquad \max_{0 \leqq x \leqq X} |g(x)| = A.$$

If $0 \leqq x < x' \leqq X$, then we have

$$|(h(x))_{1-\alpha} - (h(x'))_{1-\alpha}| \leqq A(x'-x).$$

Hence $(h(x))_{1-\alpha}$ is an absolutely continuous function.

---

[3] Another criterion for the absolute convergence of a Fourier series at a fixed point can be found in the note: K. K. Chen (2).

Since $g(t)$ is bounded, $(g(x))_\alpha$ is continuous.[4] Remembering that $0 < \alpha < 1$, we have

$$(h(x))_{1-\alpha} = \int_0^x g(t)\,dt = ((g(x))_\alpha)_{1-\alpha}.$$

It follows that the two functions $h(x)$ and $(g(x))_\alpha$ are equivalent. Hence $h(x)$ belongs to the Lebesgue class $L^p(0,X)$ for every $p > 1$. Hardy and Littlewood have proved that[5] if $p \geqq 1$, $0 < \alpha < 1$, and $h(x)$ belongs to L^p, then

$$\left[\int_0^X |(h(x))_\alpha - (h(x-\delta))_\alpha|^p\,dx\right]^{1/p} = o(\delta^\alpha),$$

as $\delta \to +0$. It can be deduced from this theorem that

$$(6) \qquad (h(x))_\alpha \in \mathrm{Lip}\,\alpha.$$

But it is better to give a direct proof for (6). We have

$$(7) \quad \Gamma(\alpha)\{(h(x'))_\alpha - (h(x))_\alpha\}$$
$$= \int_0^x \{(x'-t)^{\alpha-1} - (x-t)^{\alpha-1}\}h(t)\,dt + \int_0^{x'} (x'-t)^{\alpha-1}h(t)\,dt.$$

We may suppose that $h(t)$ is continuous, so that

$$|h(t)| \leqq M\Gamma(\alpha+1) \qquad (0 \leqq t \leqq X),$$

if M is sufficiently large. Assuming $x' > x$, it follows from (7) that

$$|(h(x'))_\alpha - (h(x))_\alpha| \leqq \alpha M \int_0^x \{(x-t)^{\alpha-1} - (x'-t)^{\alpha-1}\}dt + M(x'-x)^\alpha$$
$$= M(x'^\alpha - x^\alpha) + 2M(x'-x)^\alpha \leqq 3M(x'-x)^\alpha,$$

since $0 < \alpha < 1$. This proves (6).

Finally, by definition, $(h(x))_{-\alpha} = g(x)$. Hence

$$((h(x))_{-\alpha})_\alpha = (g(x))_\alpha.$$

Since $((g(x))_\alpha$ is equivalent to $h(x)$, the proof is completed.

4. LEMMA 2. *Suppose that* $q + 1 > \alpha > 0$, $\alpha < 1$, $0 < w \leqq \pi$, *and that*

$$Z(w) = \int_0^w u^{q-\alpha} \int_u^\pi (t-u)^{\alpha-1} t^{-q} \cos nt\,dt\,du.$$

---

[4] G. H. Hardy (1).

[5] G. H. Hardy and J. E. Littlewood (2).

Then there is a constant C such that

$$|nZ(w)| \begin{cases} \leqq C(nw)^{-\alpha} & (nw \geqq 1), \\ \leqq C(nw)^{1-\alpha} & (nw < 1,\ \alpha \leqq q), \\ \leqq C(nw)^{1-\alpha+q} & (nw < 1,\ \alpha > q). \end{cases} \tag{8}$$

Proof. In the first place, it follows from

$$\begin{aligned} Z(\pi) &= \int_0^\pi t^{-q} \cos nt \int_0^t u^{q-\alpha}(t-u)^{\alpha-1} du\, dt \\ &= B(\alpha, q-\alpha+1) \int_0^\pi \cos nt\, dt \\ &= 0 \end{aligned}$$

that

$$\begin{aligned} Z(w) &= Z(\pi) - \int_w^\pi u^{q-\alpha} \int_u^\pi (t-u)^{\alpha-1} t^{-q} \cos nt\, dt\, du \\ &= -\int_w^\pi t^{-q} \cos nt \int_w^\pi u^{q-\alpha}(t-u)^{\alpha-1} du\, dt \\ &= -\int_w^\pi \cos nt \int_{w/t}^1 v^{q-\alpha}(1-v)^{\alpha-1} dv\, dt \qquad (9) \\ &= (1/n) \int_w^\pi \sin nt \cdot (w/t^2)(w/t)^{q-\alpha}(1-w/t)^{\alpha-1} dt \\ &= (1/n) w^{1+q-\alpha} \int_w^\pi (t-w)^{\alpha-1} t^{-1-q} \sin nt\, dt. \end{aligned}$$

The second mean value theorem gives two numbers t_1 and t_2 such that

$$nw^\alpha Z(w) = \int_w^{t_1} (t-w)^{\alpha-1} \sin nt\, dt, \qquad w < t_1 < \pi,$$

and that $w < t_2 < t_1$,

$$|nw^\alpha Z(w)| \leqq \int_w^{w+1/n} (t-w)^{\alpha-1} dt + |n^{1-\alpha} \int_w^{t_2} \sin nt\, dt|.$$

Hence

$$|nw^\alpha Z(w)| \leqq (2 + 1/\alpha) n^{-\alpha}.$$

This proves (8) for the case $nw \geqq 1$.

Secondly, let $nw \leqq 1$; we have

$$\begin{aligned} & w^{1+q-\alpha} \int_w^\pi (t-w)^{\alpha-1} t^{-1-q} \sin nt\, dt \\ &= (nw)^{1+q-\alpha} \int_0^n y^{\alpha-1} \frac{\sin(y+nw)}{(y+nw)^{1+q}} dy \\ &= (nw)^{1+q-\alpha} \left(\int_0^{\frac{1}{2}\pi-1} + \int_{\frac{1}{2}\pi-1}^{n(\pi-w)} \right) \\ &= (nw)^{1+q-\alpha} \{ \lambda_n \int_0^{\frac{1}{2}\pi-1} y^{\alpha-1}(y+nw)^{-q} dy + \mu_n \int_{\frac{1}{2}\pi-1}^\infty y^{\alpha-q-2} dy \}, \end{aligned}$$

where $|\lambda_n| < 1$, $|\mu_n| < 1$. Since $q + 1 > \alpha$, the integral over $(\pi/2 - 1, \infty)$ is convergent. The integral

$$I = \int_0^{\frac{1}{2}\pi - 1} y^{\alpha-1}(y + nw)^{-q} dy$$

is $O(1)$, when $\alpha > q$. And if $\alpha \leqq q$, then

$$I < (nw)^{-q} \cdot (1/\alpha)(\pi/2)^{\alpha}.$$

This completes the proof of the lemma.

5. LEMMA 3. *If $P(t)$ is of bounded variation in $(0, \pi)$ then the series*

$$\sum_{n=1}^{\infty} n^{-1} \int_0^{1/n} (nt)^{\epsilon} \, | d\phi(t) | \quad \textit{and} \quad \sum_{n=1}^{\infty} n^{-1} \int_{1/n}^{\pi} (nt)^{-\epsilon} \, | d\phi(t) |$$

are both convergent, where $\epsilon > 0$.

Proof. The function

$$H(t) = \sum_{n=1}^{\infty} n^{-1} \min \{ (nt)^{\epsilon}, (nt)^{-\epsilon} \}$$

bounded in $(0, \pi)$. In fact,

$$H(t) = \sum_{nt \leqq 1} n^{-1}(nt)^{\epsilon} + \sum_{nt > 1} n^{-1}(nt)^{-\epsilon}.$$

This is clearly $O(1)$. The sum of any one of the two series in the lemma, is not greater than $\int_0^{\pi} H(t) \, | d\phi(t) |$ which is a finite number, since $\phi(t)$ is of bounded variation. Lemma 3 is thus proved.

6. Guided by these lemmas, we were led to the following theorem which is an extension of Theorem 1.

THEOREM 2. *If $0 < \alpha < 1$, $q \geqq \alpha$, and the function*

$$\psi(t) = t^{\alpha - q} \frac{d}{dt} \int_0^t \frac{u^q \phi(u) \, du}{(t - u)^{\alpha}} \tag{10}$$

exists and is of bounded variation in $(0, \pi)$, then the Fourier series of f converges absolutely at the point x.

This reduces to Theorem 1, when $\alpha = q = p$.

Proof. We have to deduce the absolute convergence of the series

$$\sum_{n=1}^{\infty} (2/\pi) \int_0^{\pi} \phi(t) \cos nt dt \tag{11}$$

from the fact that the function $\psi(t)$ is of bounded variation in $(0,\pi)$.[6]

In virtue of $q \geqq \alpha$, the function

$$(t^q\phi(t))_{-\alpha} = \frac{1}{\Gamma(1-\alpha)}\,\frac{d}{dt}\int_0^t \frac{u^q\phi(u)\,du}{(t-u)^\alpha}$$

being the product of $\psi(t)$ and $t^{q-\alpha}/\Gamma(1-\alpha)$, is of bounded variation in $(0,\pi)$.

It follows from Lemma 1 that the relation

$$\phi(t) = t^{-q}((t^q\phi(t))_{-\alpha})_\alpha\,.$$

holds almost everywhere. Therefore the n-th term of (11) can be written as

$$\begin{aligned}\int_0^\pi & t^{-q}\cos nt\cdot(1/\Gamma(\alpha))\int_0^t (t-u)^{\alpha-1}(u^q\phi(u))_{-\alpha}\,du\,dt\\ &= (1/\Gamma(\alpha))\int_0^\pi (u^q\phi(u))_{-\alpha}\int_u^\pi (t-u)^{\alpha-1}t^{-q}\cos nt\,dt\,du\\ &= (1/\Gamma(\alpha))\int_0^\pi u^{\alpha-q}(u^q\phi(u))_{-\alpha}\,dZ(u).\end{aligned}$$

On integration by parts, we obtain

$$\Gamma(\alpha)\Gamma(1-\alpha)\int_0^\pi \phi(t)\cos nt\,dt = -\int_0^\pi Z(w)\,d\psi(w)$$

since $Z(0) = Z(\pi) = 0$. Hence, by Lemma 2,

$$\left|\int_0^\pi \phi(t)\cos nt\,dt\right| \leqq (C/n)\int_0^{1/n}(nw)^{1-\alpha}\,|\,d\psi(w)\,| + (C/n)\int_{1/n}^\pi (nw)^{-\alpha}\,|\,d\psi(w)\,|. \tag{12}$$

The absolute convergence of (11) follows from (12) and Lemma 3. The theorem is thus proved.

7. If we equate α and q in Theorem 2, we obtain Theorem 4. In particular, if the derivative $\phi'(t)$ exists and both $\phi(t)$ and $t\phi'(t)$ are of bounded variation in $(0,\pi)$, then

$$\begin{aligned}\psi(t) &\equiv \frac{d}{dt}\int_0^t \frac{u^p\phi(u)\,du}{(t-u)^p} = \frac{d}{dt}\int_0^1 \frac{tv^p\phi(tv)\,dv}{(1-v)^p}\\ &= \int_0^1 \frac{v^p(\phi(tv) + tv\phi'(tv))\,dv}{(1-v)^p}\,.\end{aligned}$$

---

[6] Cf. L. S. Bosanquet (1).

which is evidently also of bounded variation in $(0, \pi)$. Hence we obtain the following

THEOREM 3. *If $\phi(t)$ and $t\phi'(t)$ are both of bounded variation in $(0, \pi)$ then the Fourier series of f converges absolutely at the point x.*

8. Write σ_n^α for the n-th Cesàro mean of order α of the series $\sum_{n=0}^{\infty} a_n$. Thus, writing $(\alpha)_n = \dfrac{\Gamma(n+\alpha+1)}{\Gamma(n+1)\Gamma(\alpha+1)}$ for $\alpha > -1$,

$$\sigma_n^\alpha = (1/(\alpha)_n) \sum_{\nu=0}^{\infty} (\alpha)_{n-\nu} a_\nu.$$

The series $\sum a_n$ is said to be absolutely summable (C, α) or summable $| C, \alpha |$, if the series

$$\sum_{n=0}^{\infty} | \sigma_n^\alpha - \sigma_{n-1}^\alpha |$$

with $\sigma_{-1}^\alpha = 0$, converges. In this case, letting $\sigma_n^\alpha \to s$, we write

$$\sum_{n=0}^{\infty} a_n = s \mid C, \alpha \mid,$$

In particular, summability $| C, 0 |$ is the same as absolute convergence.

LEMMA 4. *If $\alpha > -1$ and $\sum a_n = s \mid C, \alpha \mid$, then for $\epsilon > 0$,*

$$\sum a_n = s \mid C, \alpha + \epsilon \mid.$$

This important theorem is due to Kogbetliantz.[7] It seems however that the proposition is new in the case $-1 < \alpha < 0$. We give a proof here.

Proof. Without loss of generality, we can assume that $\epsilon < 1$. Writing $\alpha + \epsilon = \beta$, we have, by Abel's transformation,

$$\begin{aligned}(\beta)_n \sigma_n^\beta &= \sum_{\nu=0}^{n} (\epsilon - 1)_{n-\nu} (\alpha)_\nu \sigma_\nu^\alpha \\ &= \sum_{\nu=0}^{n-1} (\sigma_\nu^\alpha - \sigma_{\nu+1}^\alpha) \sum_{\mu=0}^{\nu} (\alpha)_\mu (\epsilon - 1)_{n-\mu} + \sigma_n^\alpha (\beta)_n.\end{aligned}$$

It follows that

$$\sigma_n^\beta - \sigma_{n-1}^\beta = ((\alpha)_n/(\beta)_n)(\sigma_n^\alpha - \sigma_{n-1}^\alpha) + \sum_{\nu=0}^{n-2} (\sigma_\nu^\alpha - \sigma_{\nu+1}^\alpha)(n, \nu),$$

---

[7] E. Kogbetliantz (1), (2).

where

$$(n, \nu) = \sum_{\mu=0}^{\nu} (\alpha)_\mu \left\{ \frac{(\epsilon - 1)_{n-\mu}}{(\beta)_n} - \frac{(\epsilon - 1)_{n-\mu-1}}{(\beta)_{n-1}} \right\} = (n, \nu)_1 + (n, \nu)_2,$$

$$(n, \nu)_1 = \frac{1 + \alpha}{1 - \epsilon} \frac{1}{(\beta)_{n-1}} \sum_{\mu=0}^{\nu} (\alpha)_\mu (\epsilon - 2)_{n-\mu},$$

$$(n, \nu)_2 = \frac{\beta(\alpha + 1)}{(\epsilon - 1)(\beta + 1)(\beta + 1)_{n-1}} \sum_{\mu=0}^{\nu} (\alpha + 1)_\mu (\epsilon - 2)_{n-\mu}.$$

If $N > 2$, then we have

$$\sum_{n=1}^{N} | \sigma_n{}^\beta - \sigma_{n-1}{}^\beta | \leqq \sum_{n=1}^{N} ((\alpha)_n/(\beta)_n) | \sigma_n{}^a - \sigma_{n-1}{}^a | + \sum_{n=2}^{N} \sum_{\nu=0}^{n-2} | ((\sigma_\nu{}^a - \sigma_{\nu+1}{}^a)(n, \nu) |.$$

The first term is $O(1)$, as $N \to \infty$, since $(\beta)_n > (\alpha)_n$. The last term is equal to

$$\sum_{\nu=0}^{N-2} | \sigma_\nu{}^a - \sigma_{\nu+1}{}^a | \sum_{n=\nu+2}^{N} | (n, \nu) |.$$

Hence it is enough to show that $\sum_{\nu+2}^{N} | (n, \nu) |$ is uniformly bounded in N and ν.

There is a constant K such that

$$\sum_{n=2\nu+1}^{N} | (n, \nu)_1 | < K \sum_{n=2\nu+1}^{N} n^{-\beta} \cdot n^{\epsilon-2} \sum_{\mu=0}^{\nu} (\alpha)_\mu.$$

This is equal to

$$K(\alpha + 1)_\nu \sum_{n=2\nu+1}^{N} n^{-2-a} = O(1),$$

uniformly in N and ν.

If $\nu + 2 \leqq n \leqq 2\nu$, then $(\beta)_{n-1} = O(\nu^\beta)$ and

$$\begin{aligned} \sum_{n=\nu+2}^{2\nu} \sum_{u=0}^{\nu} (\alpha)_\mu | (\epsilon - 2)_{n-\mu} | &= - \sum_{n=\nu+2}^{2\nu} \sum_{\mu=0}^{\nu} (\alpha)_\mu (\epsilon - 2)_{n-\mu} \\ &= - \sum_{\mu=0}^{\nu} (\alpha)_\mu \sum_{n=\nu+2}^{2\nu} (\epsilon - 2)_{n-\mu} \\ &= \sum_{\mu=0}^{\nu} (\alpha)_\mu ((\epsilon - 1)_{\nu-\mu+1} - (\epsilon - 1)_{2\nu-\mu}). \end{aligned}$$

This is less than $\sum_{\mu=0}^{\nu} (\alpha)_\mu (\epsilon - 1)_{\nu-\mu} = (\beta)_\nu$. Hence

$$\sum_{n=\nu+2}^{2\nu} (1/(\beta)_{n-1}) \sum_{\mu=0}^{\nu} (\alpha)_\mu (\epsilon - 2)_{n-\mu} = O(\nu^{-\beta})(\beta)_\nu = O(1).$$

Therefore we have

$$\sum_{n=\nu+1}^{\infty} |(n,\nu)_1| = O(1).$$

In a similar manner, we can prove

$$\sum_{n=\nu+1}^{\infty} |(n,\nu)_2| = O(1).$$

The proposition is thus established.

9. It has now been shewn that the absolute summability of negative order is a property which is stronger than absolute convergence. Let us next consider the absolute summability of negative order. For this purpose, we write

$$[\chi(t)]_\alpha = \Gamma(1+\alpha)t^{-\alpha}(\chi(t))_\alpha$$

as the mean function of order α of $\chi(t)$. Thus

$$[\chi(t)]_0 = \chi(t),$$
$$[\chi(t)]_\alpha = (\alpha/t^\alpha)\int_0^t (t-u)^{\alpha-1}\chi(u)\,du \qquad (\alpha > 0,\ t > 0),$$

provided that $\chi(t)$ is integrable in the Lebesgue sense over $(0, t_0)$. The latter equation in the case $0 < \alpha < 1$ holds good for almost all values of t in $(0, t_0)$.[8] The mean functions of negative order exist with additional hypotheses.

THEOREM 4. *Suppose that* $-1 < -\alpha < 0$, $q \geqq \alpha$ *and that at the point* x *the function*

$$t^{-q}[t^q\phi(t)]_{-\alpha} \tag{13}$$

is of bounded variation in $(0,\pi)$; *then, for* $\epsilon > 0$, *the Fourier series of* f *is summable* $|C, -\alpha+\epsilon|$.

Before giving the proof of this theorem it is convenient to prove the following three lemmas.

10. LEMMA 5. *Write* $g^\alpha(n,t)$ *for the* n*-th Cesàro mean of order* α *of* $2\pi^{-1}\sin nt$. *If* $-1 < \beta < 0$,[9] *then*

$$|g^\beta(n,t)| \leqq Ant(1+nt)^{-1-\beta}, \qquad (t > 0).$$

---

[8] G. H. Hardy (1).

[9] If $\alpha \geqq 0$, then $g^\alpha(n,t) = O(1+nt)^{-\alpha} + O(1+nt)$. See Obreschkoff (1).

Proof. In fact, we have

$$g^{\beta}(n,t) = \frac{2}{\pi}\,\frac{1}{(\beta)_n}\sum_{\nu=0}^{n}(\beta-1)_{n-\nu}\sin\nu t.$$

Abel's transformation gives

$$\begin{aligned}\sum_{\nu=0}^{n}(\beta-1)_{n-\nu}\sin\nu t &= \sum_{\nu=0}^{n-1}(\beta)_{\nu}\Delta\sin(n-\nu)t\\ &= -2\sin(t/2)\sum_{\nu=0}^{n-1}(\beta)_{\nu}\cos(n-\nu-\tfrac{1}{2})t\end{aligned}$$

which is numerically less than a constant multiple of $tn^{1-\beta}$. This proves the lemma for the case $nt \leqq 1$. To complete the proof, we rewrite

$$\sum_{\nu=0}^{n}(\beta-1)_{n-\nu}\sin\nu t = \Big(\sum_{\nu t\leqq 1}+\sum_{nt>1}\Big)(\beta-1)_{\nu}\sin(n-\nu)t.$$

Now, if $nt > 1$,

$$\Big|\sum_{\nu t>1}(\beta-1)_{\nu}\sin(n-\nu)t\Big| \leqq \sum_{\nu>1/t}^{\infty}|(\beta-1)_{\nu}| \leqq At^{-\beta},$$

$$\begin{aligned}\sum_{\nu t\leqq 1}(\beta-1)_{\nu}\sin(n-\nu)t &= -2\sin(t/2)\sum_{\nu t\leqq 1}(\beta)_{\nu}\cos(n-\nu-\tfrac{1}{2})t + O(t^{-\beta})\\ &= O(t^{-\beta}).\end{aligned}$$

Therefore we obtain

$$g^{\beta}(n,t) = O(nt)^{-\beta}.$$

The lemma is thus proved.

From the proof of Lemma 5, we can state the following result.

LEMMA 6. *If* $-1 < \beta < 0$, $nt \geqq 1$, *then the relation*

$$(1/(\beta)_n)\sum_{\nu t\leqq 1}^{\mu}(\beta-1)_{\nu}\sin(n-\nu)t = O(nt)^{-\beta}$$

holds uniformly for $t^{-1} \leqq \mu \leqq n$.

11. LEMMA 7. *Suppose that* $-1 < -\alpha < \beta < 0$, $1 + q > \alpha - \beta$ *and that*

$$Z_{\beta}(w) = \int_{0}^{w}u^{q-\alpha}\int_{u}^{\pi}(t-u)^{\alpha-1}t^{-q}(d/dt)\,g^{\beta}(n,t)\,dt\,du. \tag{14}$$

Then for $0 < w \leqq \pi$,

$$|Z_{\beta}(w)| \leqq (nw)^{-\beta}(1+nw)^{-\alpha}. \tag{15}$$

Proof. As in the proof of Lemma 2, we have

$$Z_\beta(\pi) = B(\alpha, q-\alpha+1) \int_0^\pi (d/dt) g^\beta(n,t)\,dt = 0.$$

Hence

$$(16) \qquad Z_\beta(w) = -\int_w^\pi (d/dt) g^\beta(n,t) \int_{w/t}^1 v^{q-\alpha}(1-v)^{\alpha-1} dv\,dt$$
$$= w^{1+q-\alpha} \int_w^\pi (t-w)^{\alpha-1} t^{-1-q} g^\beta(n,t)\,dt.$$

Write $g^\beta(n,t) = g_1(n,t) + g_2(n,t)$ with

$$(\beta)_n g_1(n,t) = \sum_{\nu=0}^{[n/2]} (\beta-1)_\nu \sin(n-\nu)t.$$

Firstly, assuming $nw \geqq 2$, let us consider the following two integrals

$$Z_{\beta}{}^{j}(w) = w^{1+q-\alpha} \int_w^\pi (t-w)^{\alpha-1} t^{-1-q} g_j(n,t)\,dt; \qquad (j=1,2).$$

By Lemma 6, we have

$$(17) \qquad w^{1+q-\alpha} \int_w^{w+1/n} (t-w)^{\alpha-1} t^{-1-q} g_1(n,t)\,dt$$
$$= w^{1+q-\beta} O(nw)^{-\beta} \int_w^{w+1/n} (t-w)^{\alpha-1} t^{-1-q}\,dt = O(nw)^{-\alpha-\beta}.$$

The second mean value theorem gives a number w_1 for which $w < w_1 < 2w$, such that

$$w^{1+q-\alpha} \int_{w+1/n}^{2w} (t-w)^{\alpha-1} t^{-1-q} g_1(n,t)\,dt = O(n)(nw)^{-\alpha} \int_{w+1/n}^{w_1} g_1(n,t)\,dt.$$

We have

$$\int_{w+1/n}^{w_1} g_1(n,t)\,dt = \frac{1}{(\beta)_n} \left[\sum_{\nu=0}^{n/2} (\beta-1)_\nu \frac{-\cos(n-\nu)t}{n-\nu} \right]_{w+1/n}^{w_1}.$$

Since $t^{-1} < n/2$, we have

$$\sum_{\nu t \leqq 1} (\beta-1)_\nu \frac{\cos(n-\nu)t}{n-\nu} = \sum_{\nu t > 1} (\beta)_\nu \Delta \frac{\cos(n-\nu)t}{n-\nu} + O(n^{-1}t^{-\beta}).$$

And in virtue of $\beta < 0$, we have

$$\sum_{\nu t > 1} (\beta-1)_\nu \frac{\cos(n-\nu)t}{n-\nu} = O(n^{-1}t^{-\beta}).$$

Therefore we obtain

$$(18) \qquad w^{1+q-\alpha} \int_{w+1/n}^{2w} (t-w)^{\alpha-1} t^{-1-q} g_1(n,t)\,dt = O(nw)^{-\alpha-\beta}.$$

Integrating by parts, we have

$$\int_{2w}^{\pi} (t-w)^{\alpha-1} t^{-1-q} g_1(t)\,dt$$

$$= \left[(t-w)^{\alpha-1} t^{-1-q} \sum_{\nu=0}^{n/2} \frac{(\beta-1)_\nu}{(\beta)_n} \frac{-\cos(n-\nu)t}{n-\nu} \right]_{2w}^{\pi}$$

$$+ \int_{2w}^{\pi} \frac{d}{dt} \{ (t-w)^{\alpha-1} t^{-1-q} \} \sum_{\nu=0}^{n/2} \frac{(\beta-1)_\nu}{(\beta)_n} \frac{\cos(n-\nu)t}{n-\nu}\,dt$$

$$= O(n^{-1-\beta}) + O(w^{\alpha-q-2} n^{-1-\beta}) + \int_{2w}^{\pi} - \frac{d}{dt} \{ (t-w)^{\alpha-1} t^{-1-q} \} O(n^{-1}) (nt)^{-\beta} dt,$$

since, for $t > w$,

$$-(d/dt)\{ (t-w)^{\alpha-1} t^{-1-q} \} = t^{-2-q} (t-w)^{\alpha-2} \{ (2+q-\alpha)t - (1+q)w \}$$
$$> t^{-2-q} (t-w)^{\alpha-2} (1-\alpha) w > 0.$$

Now, we have

$$0 < \int_{2w}^{\pi} - (d/dt) \{ (t-w)^{\alpha-1} t^{-1-q} \} \cdot t^{-\beta} dt$$
$$= O(1) + w^{\alpha-q-\beta-2} - \beta \int_{2w}^{\pi} (t-w)^{\alpha-1} t^{-2-q-\beta} dt = O(w^{\alpha-q-\beta-2}),$$

since $q + 2 > \alpha - \beta$. It follows that

$$(19) \qquad w^{1+q-\alpha} \int_{2w}^{\pi} (t-w)^{\alpha-1} t^{-1-q} g_1(t)\,dt = O(nw)^{-1-\beta}.$$

Collecting (17), (18) and (19), we obtain

$$(20) \qquad Z_\beta^1(w) = O(nw)^{-\alpha-\beta}, \qquad (nw \geqq 1).$$

We are now in a position to consider $Z_\beta^2(w)$. We have

$$Z_\beta^2(w) = w^{1+q-\alpha} \int_{w}^{\pi} (t-w)^{\alpha-1} t^{-1-q} g_2(n,t)\,dt$$
$$= \int_0^w u^{q-\alpha} \int_u^{\pi} (t-u)^{\alpha-1} t^{-q} (d/dt) g_2(n,t)\,dt\,du,$$

since the function $(\beta)_n g_2(n,t) = \sum_{2\nu < n} (\beta-1)_{n-\nu} \sin \nu t$ vanishes at 0 and π. Therefore we may write

$$(21) \qquad Z_\beta^2(w) = \int_w^{\pi} t^{-q} (dg_2/dt) \int_0^w u^{q-\alpha} (t-u)^{\alpha-1} du\,dt$$
$$+ \int_0^w t^{-q} (dg_2/dt) \int_0^t u^{q-\alpha} (t-u)^{\alpha-1} du\,dt.$$

The last term is a constant multiple of

$$(22) \quad \int_0^w (dg_2/dt)\,dt = g_2(n, w) = (1/(\beta)_n) \sum_{2\nu < n} (\beta - 1)_{n-\nu} \sin \nu w$$
$$= (1/(\beta)_n)\, O(n^{\beta-1}) \max_{\mu} |\sin w + \sin 2w + \cdots + \sin \mu w|$$
$$= O(nw)^{-1},$$

since $(\beta - 1)_{n-\nu}$ is increasing in ν. And it follows from the second mean value theorem that

$$(23) \quad \int_w^\pi (dg_2/dt) \int_0^{w/t} v^{q-\alpha}(1 - v)^{\alpha-1} dv dt = B(\alpha, 1 + q - \alpha) \int_w^b (dg_2/dt)\,dt$$
$$= O(nw)^{-1}.$$

Combining the results (20), (21), (22) and (23), we obtain

$$(24) \qquad Z_\beta(w) = Z_\beta{}^1(w) + Z_\beta{}^2(w) = O(nw)^{-\alpha-\beta}, \qquad (nw \geqq 2).$$

Secondly, let $nw < 2$. Using Lemma 5, we have

$$(25) \quad w^{1+q-\alpha} \int_w^{2w} (t - w)^{\alpha-1} t^{-1-q} g^\beta(n, t)\,dt = w^{1+q-\alpha} \int_w^{2w} (t - w)^{\alpha-1} t^{-1-q} O(nw)\,dt$$
$$= w^{-\alpha} \int_w^{2w} (t - w)^{\alpha-1} O(nw)\,dt = O(nw).$$

By Lemma 5, $g^\beta(n, t) = O(nt)^{-\beta}$ for $t > 0$. It follows that

$$(26) \quad w^{1+q-\alpha} \int_{2w}^{\pi} (t - w)^{\alpha-1} t^{-1-q} g^\beta(n, t)\,dt = O(n^{-\beta})\, w^{1+q-\alpha} \int_{2w}^{\pi} (t - w)^{\alpha-1} t^{-1-q-\beta} dt$$
$$= O(nw)^{-\beta} \int_2^{\pi/w} (v - 1)^{\alpha-1} v^{-1-q-\beta} dv = O(nw)^{-\beta},$$

since the integral $\int_2^\infty (v - 1)^{\alpha-1} v^{-1-q-\beta} dv$ converges, by the hypothesis $\alpha - \beta < 1 + q$. From (25) and (26) we have

$$(27) \qquad Z_\beta(w) = O(nw)^{-\beta} \qquad (nw \leqq 2).$$

The two relations (24) and (27) establishes the lemma.

12. Proof of Theorem 4. By hypothesis, the functions

$$\psi(t) = t^{-q}[t^q \phi(t)]_{-\alpha} = \frac{1}{\Gamma(1 - \alpha)}\, t^{\alpha-q} \frac{d}{dt} \int_0^t \frac{u^q \phi(u)\,du}{(t - u)^\alpha}$$

and $t^{q-\alpha}$ are of bounded variation in $(0, \pi)$. Hence the product

$$t^{q-\alpha}\psi(t)=\frac{d}{dt}\int_0^t\frac{u\phi(u)\,du}{(t-u)^\alpha}=\Gamma(1-\alpha)\,(t^q\phi(t))_{-\alpha}$$

is bounded in $(0,\pi)$. It follows from Lemma 1 that

$$\phi(t)=((t^q\phi(t))_{-\alpha})_\alpha t^{-q}, \tag{28}$$

except for a null set of points t.

Write $r_n{}^\beta$ and $\sigma_n{}^\beta$ for the n-th Cesàro mean of order β of

$$(2/\pi)\int_0^\pi\phi(t)\,(d/dt)\sin nt\,dt \quad\text{and}\quad (2/\pi)\int_0^\pi\phi(t)\cos nt\,dt$$

respectively. An easy calculation gives

$$\sigma_n{}^\beta-\sigma_{n-1}{}^\beta=n^{-1}r_n{}^\beta.$$

Assuming $-1<-\alpha<\beta<0$. We have to show that $\Sigma\mid n^{-1}r_n{}^\beta\mid$ converges. By (28), we have

$$r_n{}^\beta=\int_0^\pi((t^q\phi(t))_{-\alpha})_\alpha t^q\,\frac{dg^\beta(n,t)}{dt}\,dt.$$

As in the proof of Theorem 2, this can be rewritten as

$$r_n{}^\beta=\frac{1}{\Gamma(\alpha)}\int_0^\pi w^{\alpha-q}(w^q\phi(w))_{-\alpha}dZ_\beta(w)=-\frac{\sin\alpha\pi}{\pi}\int_0^\pi Z_\beta(w)\,d\psi(w).$$

It follows from Lemma 7 that

$$\mid r_n{}^\beta/n\mid\leqq An^{-1}\int_0^{1/n}(nw)^{-\beta}\mid d\psi(w)\mid+An^{-1}\int_{1/n}^\pi(nw)^{-\alpha-\beta}\mid d\psi(w)\mid.$$

On account of $-\beta>0$ and $-\alpha-\beta<0$, the convergence of the series $\Sigma\mid n^{-1}r_n{}^\beta\mid$ follows from Lemma 3. This establishes the theorem.

Putting $\alpha=q=p$, it has been shown that the function

$$\frac{d}{dt}\int_0^t\frac{u^p\phi(u)\,du}{(t-u)^p}\qquad(0<p<1)$$

is of bounded variation in $(0,\pi)$, if both $\phi(t)$ and $t\phi'(t)$ are of bounded variation in $(0,\pi)$. It follows from Theorem 4 that the series $\Sigma\mid n^{-1}r_n{}^\beta\mid$ converges for $\beta>-p$. Hence the Fourier series is summable $\mid C,\alpha\mid$ for every $\alpha>-1$. We have now obtained the following

THEOREM 5. *If $\phi(t)$ and $t\phi'(t)$ are both of bounded variation in $(0, \pi)$ then the Fourier series of f, at the point x, is absolutely summable (C, α) for every $\alpha > -1$.*

NATIONAL UNIVERSITY OF CHEKIANG,
MEITAN, KWEICHOW, CHINA.

REFERENCES.

L. S. Bosanquet:
- (1) "The absolute Cesàro summability of a Fourier series," *Proceedings of the London Mathematical Society* (2), vol. 41 (1936), pp. 517-528.

K. K. Chen:
- (1) "On the class of functions with absolutely convergent Fourier series," *Proc. Imp. Acad.*, vol. 4 (1928), pp. 348-351.
- (2) "On the absolute Cesàro summability of negative order for a Fourier series at a given point," *American Journal of Mathematics*, vol. 66 (1944), pp. 299-312.

G. H. Hardy:
- (1) "Notes on some points in the integral calculus (XLVIII): On some properties of integrals of fractional order," *Messenger of Mathematics*, vol. 47 (1918), pp. 145-150.

G. H. Hardy and J. E. Littlewood:
- (1) "Notes on the theory of series (IX): On the absolute convergence of Fourier series," *Journal of the London Mathematical Society*, vol. 3 (1928), pp. 250-253.
- (2) "Some properties of fractional integrals (I)," *Mathematische Zeitschrift*, vol. 27 (1928), pp. 565-606.

E. Kogbetliantz:
- (1) "Sur les séries absolument sommables par la méthode des moyennes arithmétiques," *Bull. des Sci. Math.* (2), vol. 49 (1925), pp. 234-256.
- (2) *Mémorial des Sci. Math.*, vol. 51: "Sommation des séries et intégrales divergentes par les moyennes arithmétiques et typiques," (1931), Théorème VIII.

N. Obreschkoff:
- (1) "Sur la sommation des séries trigonométriques de Fourier par les moyennes arithmétiques," *Bull. de la Soc. Math. de France*, vol. 62 (1934), pp. 84-109 and 167-184.

A. Zygmund:
- (1) *Trigonometrical series* (1935).

THE RADICAL AND SEMI-SIMPLICITY FOR ARBITRARY RINGS.*[1]

By N. JACOBSON.

The radical of an algebra with a finite basis, or, more generally, of a ring $\mathfrak{A}$ that satisfies the descending chain condition is defined to be the join of the nil right (left) ideals of $\mathfrak{A}$. The importance of the radical for the structure theory of these rings is due to the facts that 1) the radical $\mathfrak{R}$ is a two-sided ideal whose difference ring $\mathfrak{A} - \mathfrak{R}$ is semi-simple in the sense that its radical is 0, and 2) the structure of semi-simple rings satisfying the descending chain condition can be subjected to a thorough analysis that leads in many important cases to a complete classification. Several investigations of nil ideals in arbitrary rings have been made recently but none of these has led to a structure theory for general semi-simple rings.[2] This is one of a number of indications that in order to develop a satisfactory structure theory for arbitrary rings it is necessary to abandon the concept of a nil ideal in defining the radical.

Other possibilities for defining a radical are afforded by two important characterizations of the radical $\mathfrak{R}$ of an algebra $\mathfrak{A}$ with a finite basis. One of these, due to Perlis, makes use of the notion of quasi-regularity.[3] An element z of $\mathfrak{A}$ is right quasi-regular if there exists a z' in $\mathfrak{A}$ such that $z + z' + zz' = 0$. Perlis has shown that $z \in \mathfrak{R}$ if and only if $u + z$ is right quasi-regular for all right quasi-regular u. A second characterization of $\mathfrak{R}$ for algebras with an identity is that $\mathfrak{R}$ is the intersection of the maximal right (left) ideals of $\mathfrak{A}$.[4] A start in the investigation of the first characterization as a possibility for defining a radical for an arbitrary ring $\mathfrak{A}$ was made by Baer, who showed that the totality $\mathfrak{R}$ of elements z that generate right ideals containing only right quasi-regular elements is a right ideal.[5]

The point of departure of the present investigation is the observation that $\mathfrak{R}$ is a two-sided ideal and that $\mathfrak{R}$ coincides with the two-sided ideal

---

* Received September 6, 1944; Revised January 5, 1945.

[1] Presented to the Society October 28, 1944.

[2] Baer [1] and Levitzki [1] and [2].

[3] Perlis [1].

[4] Jacobson [1], p. 66. Cf. also G. Birkhoff [1].

[5] Baer [1], p. 562. This definition of the radical has been independently proposed by Hille and Zorn who proved that $\mathfrak{R}$ is a two-sided ideal and that if $\mathfrak{A}$ has an identity, $\mathfrak{R}$ is the intersection of the maximal right (left) ideals of $\mathfrak{A}$. These results were announced by Professor Hille in his Colloquium Lectures in August 1944.

defined in a similar manner using left quasi-regularity and left ideals. We define the radical of $\mathfrak{A}$ to be the ideal $\mathfrak{R}$. In the first part of the paper we establish a number of "radical-like" properties of $\mathfrak{R}$ and we investigate $\mathfrak{R}$ in several important special cases. For rings that satisfy the descending chain condition $\mathfrak{R} = \mathfrak{N}$ where $\mathfrak{N}$ is defined as above. If $\mathfrak{A}$ is a commutative normed ring $\mathfrak{R}$ coincides with the radical as defined by Gelfand to be the totality of generalized nilpotent elements.[6] A similar characterization holds for non-commutative normed rings.

In the latter half of the paper we investigate the relation between $\mathfrak{R}$ and the intersection of the maximal right ideals of $\mathfrak{A}$. We show that if $\mathfrak{A} \neq \mathfrak{R}$ then $\mathfrak{R} \geqq \Pi\mathfrak{J}$ for the maximal right ideals $\mathfrak{J}$ of $\mathfrak{A}$ and $\Pi\mathfrak{J} \geqq \mathfrak{A}\mathfrak{R}$. If $\mathfrak{A}$ has an identity, $\mathfrak{R} = \Pi\mathfrak{J}$. For arbitrary rings that contain maximal right ideals we show that $\mathfrak{R}$ is the intersection of certain two-sided ideals $\mathfrak{B}$ whose difference rings are of a special type called primitive.

This result implies that any semi-simple ring is isomorphic to a subring of the complete direct sum of primitive rings. To a certain extent this reduces the study of semi-simple rings to that of primitive rings. A tool for studying the latter is a representation theorem for these rings that states that a primitive ring is isomorphic to a dense ring of linear transformations in a vector space over a division ring. These theorems are analogues of the fundamental Wedderburn-Artin structure theorems for semi-simple rings that satisfy the descending chain condition. The Wedderburn-Artin theorems can be deduced quite simply from our results. Moreover, our results contain as special cases Stone's theorem on Boolean rings and other known results on the representability of rings as subrings of direct sums of fields.[7] We can also obtain from our theory a theory of algebraic algebras that is fairly conclusive for algebras with elements of bounded degree. These results will be published in a subsequent paper.

I am indebted to R. Baer for a number of important suggestions that led to simplifications of some of the proofs and to extensions of several of the theorems from rings with an identity to arbitrary rings.

1. Definition of the radical. Let $\mathfrak{A}$ be a ring with an identity and let z be an element of $\mathfrak{A}$ such that $1 + z$ has a right inverse u. We write $u = 1 + z'$. Then $(1 + z)(1 + z') = 1$ implies

$$z + z' + zz' = 0. \tag{1}$$

Conversely if z is an element for which there exists an element z' satisfying (1) then $1 + z$ has the right inverse $1 + z'$. This leads to

---

[7] Stone [1], McCoy and Montgomery [1], McCoy [1] and [2].

[6] Gelfand [1], p. 10.

Definition 1. An element z of an arbitrary ring $\mathfrak{A}$ is *right quasi-regular* if there exists an element z' in $\mathfrak{A}$ such that $z + z' + zz' = 0$. The element z satisfying this equation is called a *right quasi-inverse* of z.

We have noted that if $\mathfrak{A}$ has an identity then z is right quasi-regular with right inverse z' if, and only if, $1 + z$ has the right inverse $1 + z'$.

If z is any element of $\mathfrak{A}$ the totality $\{x + zx\}$ of elements $x + zx$ where x ranges over $\mathfrak{A}$ is a right ideal. If z is right quasi-regular with right quasi-inverse z' then $-z = z' + zz' \epsilon \{x + zx\}$. Hence $zx \epsilon \{x + zx\}$ and $x \epsilon \{x + zx\}$. Then $\{x + zx\} = \mathfrak{A}$. On the other hand if $\{x + zx\} = \mathfrak{A}$ then $-z = z' + zz'$ for a suitable z' and z is right quasi-regular. Hence we have the alternative

Definition 1'. An element z of a ring $\mathfrak{A}$ is *right quasi-regular* if the totality of elements $\{x + zx\} = \mathfrak{A}$.[8]

A right ideal $\mathfrak{I}$ will be called *quasi-regular* if all the elements of $\mathfrak{I}$ are right quasi-regular. Let $\mathfrak{I}_1$ and $\mathfrak{I}_2$ be quasi-regular right ideals and let $z_i \epsilon \mathfrak{I}_i$, $i = 1, 2$. There exists an element z'_1 such that $z_1 + z'_1 + z_1 z'_1 = 0$. Since $z_2 + z_2 z'_1 \epsilon \mathfrak{I}_2$ this element has a right quasi-inverse w' such that

$$(z_2 + z_2 z'_1) + w' + (z_2 + z_2 z'_1) w' = 0.$$

Hence

$$\begin{aligned}(z_1 + z_2) + (z'_1 + w' + z'_1 w') + (z_1 + z_2)(z'_1 + w' + z'_1 w') \\ = [z_1 + z'_1 + z_1 z'_1] + [(z_2 + z_2 z'_1) + w' + (z_2 + z_2 z')w'] \\ + [(z_1 + z'_1 + z_1 z'_1) w'] = 0.\end{aligned}$$

Thus $z'_1 + w' + z'_1 w'$ is a right quasi-inverse of $z_1 + z_2$. Hence $\mathfrak{I}_1 + \mathfrak{I}_2$ is quasi-regular.

Now let $\mathfrak{R}$ be the join of all the quasi-regular right ideals of $\mathfrak{A}$. Since the right ideal generated by an element z is the totality of elements $zi + za$ where i is an integer and $a \epsilon \mathfrak{A}$, it is clear that $\mathfrak{R}$ is the totality of elements z such that $zi + za$ is right quasi-regular for all integral i and all a in $\mathfrak{A}$. The above result shows that $\mathfrak{R}$ is a right ideal. We wish to show that $\mathfrak{R}$ is a two-sided ideal. Let $z \epsilon \mathfrak{R}$ and $b \epsilon \mathfrak{A}$. Then $zb \epsilon \mathfrak{R}$ and there exists an element w' such that $zb + w' + (zb)w' = 0$. Then

$$bz + (-bz - bw'z) + bz(-bz - bw'z) = -b(w' + zb + zbw')z = 0$$

and so $-(bz + bw'z)$ is a right quasi-inverse for bz. Similarly if i is an integer and $a \epsilon \mathfrak{A}$, $(bz)i + (bz)a = b(zi + za)$ is right quasi-regular. Hence $bz \epsilon \mathfrak{R}$ and we have proved

---

[8] This definition of quasi-regularity is due to Baer.

THEOREM 1. *If $\mathfrak{A}$ is an arbitrary ring the join $\mathfrak{R}$ of all the quasi-regular right ideals of $\mathfrak{A}$ is a (right) quasi-regular two-sided ideal.*

Definition 2. The *radical* of a ring is the join of all the quasi-regular right ideals of the ring.

We have seen that $\mathfrak{R}$ is the set of elements z such that $zi + za$ is right quasi-regular for all integers i and all a in $\mathfrak{A}$. If $\mathfrak{A}$ is a ring with an identity, the connection between quasi-regularity and regularity shows that $\mathfrak{R}$ is the totality of elements z such that $1 + za$ has a right inverse for every a in $\mathfrak{A}$.

A similar discussion holds for left ideals. We define *left quasi-regularity, left quasi-inverse, quasi-regular left ideal* and left radical $\mathfrak{R}'$ in a manner analogous to the above. As for ordinary inverses we say that an element is *quasi-regular* if it is both right and left quasi-regular.

LEMMA 1. *If z is quasi-regular any right (left) quasi-inverse is a left (right) quasi-inverse, is uniquely determined and commutes with z.*

Let z' be a right quasi-inverse and z'' a left quasi-inverse. Then

$$z + z' + zz' = 0 \quad \text{and} \quad z + z'' + z''z = 0.$$

Hence

$$\begin{aligned} z'' &= z'' + (z + z' + zz') + z''(z + z' + zz') \\ &= z' + (z + z'' + z''z) + (z + z'' + z''z)z' = z'. \end{aligned}$$

This proves the first two statements. Since

$$z + z' + zz' = 0 = z + z' + z'z, \qquad zz' = z'z.$$

We call z' *the quasi-inverse* of z.

Now let $z \in \mathfrak{R}$. Then z has a right quasi-inverse $z' = -z - zz'$. Since $\mathfrak{R}$ is an ideal $z' \in \mathfrak{R}$. Hence z' has a right quasi-inverse. Since z is a left quasi-inverse of z' it follows by Lemma 1 that z' is quasi-regular and z is its quasi-inverse. Hence z is quasi-regular. Since $\mathfrak{R}$ is a left ideal $\mathfrak{R} \leq$ the left radical $\mathfrak{R}'$. By symmetry $\mathfrak{R}' \leq \mathfrak{R}$. This proves

THEOREM 2. *The radical of a ring $\mathfrak{A}$ is the join of all the quasi-regular left ideals of $\mathfrak{A}$.*

By the Lemma the elements of $\mathfrak{R}$ are quasi-regular.

2. Elementary properties of the radical. If $\mathfrak{A}$ is an arbitrary ring we know that we can imbed $\mathfrak{A}$ in a ring with an identity $\mathfrak{A}^*$ such that $\mathfrak{A}^* = \mathfrak{A} + (1)$, $\mathfrak{A} \wedge (1) = 0$ where (1), the ring generated by 1, is isomorphic to the ring of integers.[9] At first we assume only that $\mathfrak{A}^* = \mathfrak{A} + (1)$. Let $\mathfrak{R}(\mathfrak{A}^*)$ be the radical of $\mathfrak{A}^*$ and let $z \in \mathfrak{R}(\mathfrak{A}^*) \wedge \mathfrak{A}$. Then z has a quasi-inverse z' in $\mathfrak{A}^*$. Since $z' = -z - zz'$, $z' \in \mathfrak{A}$ and z is quasi-regular in $\mathfrak{A}$. Hence $(\mathfrak{R}(\mathfrak{A}^*) \wedge \mathfrak{A}) \leq \mathfrak{R}(\mathfrak{A})$. Since $\mathfrak{A}^* = \mathfrak{A} + (1)$ any right ideal of $\mathfrak{A}$ is a right ideal of $\mathfrak{A}^*$. Hence $\mathfrak{R}(\mathfrak{A}) \leq \mathfrak{R}(\mathfrak{A}^*)$. Thus $\mathfrak{R}(\mathfrak{A}) = \mathfrak{R}(\mathfrak{A}^*) \wedge \mathfrak{A}$.

Suppose now that $\mathfrak{A} \wedge (1) = 0$ and that (1) is isomorphic to the ring of integers. Then if $z^* \in \mathfrak{R}(\mathfrak{A}^*)$ the coset $\bar{z}^*$ of z^* in the difference ring $\mathfrak{A}^* - \mathfrak{A}$ is in the radical of this ring. Since the radical of the ring of integers is 0, $\bar{z}^* = 0$ and $z^* \in \mathfrak{A}$. Hence $\mathfrak{A} \geq \mathfrak{R}(\mathfrak{A}^*)$. From the equation $\mathfrak{R}(\mathfrak{A}) = \mathfrak{R}(\mathfrak{A}^*) \wedge \mathfrak{A}$ we obtain $\mathfrak{R}(\mathfrak{A}) = \mathfrak{R}(\mathfrak{A}^*)$.

THEOREM 3. *Let $\mathfrak{A}$ be an arbitrary ring and let $\mathfrak{A}^*$ be a ring with an identity containing $\mathfrak{A}$ such that $\mathfrak{A}^* = \mathfrak{A} + (1)$. Then the radical $\mathfrak{R}(\mathfrak{A}) = \mathfrak{R}(\mathfrak{A}^*) \wedge \mathfrak{A}$. If in addition $\mathfrak{A} \wedge (1) = 0$ and (1) is isomorphic to the ring of integers then $\mathfrak{R}(\mathfrak{A}) = \mathfrak{R}(\mathfrak{A}^*)$.*

We shall call $\mathfrak{A}$ a *radical ring* if $\mathfrak{A} = \mathfrak{R}$. If $\mathfrak{R} = 0$, $\mathfrak{A}$ is *semi-simple.* For many problems the following theorem effects a reduction to the consideration of these two extreme types of rings.

THEOREM 4. *If $\mathfrak{R}$ is the radical of $\mathfrak{A}$, $\bar{\mathfrak{A}} = \mathfrak{A} - \mathfrak{R}$ is semi-simple.*

Let $\bar{z}$ be an element of the radical of $\bar{\mathfrak{A}}$ and let z be an element in the coset $\bar{z}$. Then there exists an element z' such that $z + z' + zz' = u \in \mathfrak{R}$. Also there exists a u' such that $u + u' + uu' = 0$. Hence

$$\begin{aligned} 0 &= (z + z' + zz') + u' + (z + z' + zz')u' \\ &= z + (z' + u' + z'u') + z(z' + u' + z'u'). \end{aligned}$$

Thus z is right quasi-regular. Since the totality of elements z in the cosets $\bar{z}$ of $\mathfrak{R}(\bar{\mathfrak{A}})$ is an ideal, this totality is a quasi-regular ideal. Hence $z \in \mathfrak{R}$ and $\bar{z} = 0$.

If z is a nilpotent element of index n, $z' = \sum_1^{n-1} (-1)^i z^i$ is a quasi-inverse of z. Hence we have the following

THEOREM 5. *The radical of a ring contains every nil right (left) ideal of the ring.*

As a consequence of Theorems 4 and 5 we may prove the

COROLLARY. *If z is an element such that $\mathfrak{A} z \mathfrak{A} \leq \mathfrak{R}$ then $z \in \mathfrak{R}$.*

---

[9] Albert [1], p. 22.

For if $\mathfrak{I}$ is the right ideal generated by z, $\mathfrak{I}^3 \leq \mathfrak{A}z\mathfrak{A} \leq \mathfrak{R}$. Hence $\bar{\mathfrak{I}} = (\mathfrak{I} + \mathfrak{R}) - \mathfrak{R}$ is nilpotent in the semi-simple ring $\bar{\mathfrak{A}} = \mathfrak{A} - \mathfrak{R}$. Hence $\bar{\mathfrak{I}} = 0$, $\mathfrak{I} \leq \mathfrak{R}$ and $z \in \mathfrak{R}$.

This corollary implies that $z \in \mathfrak{R}$ if and only if za (az) is right (left) quasi-regular for all a in $\mathfrak{A}$.

The elements of the radical need not be nilpotent. This can be seen in the example: $\mathfrak{A}$ the ring of p-adic integers. By applying the definition directly or by using Corollary 2 to Theorem 18 we can show that $\mathfrak{R} = p\mathfrak{A}$. However, no element of $\mathfrak{A}$ is nilpotent. Some information on the nature of the elements of $\mathfrak{R}$ can be obtained from the following

THEOREM 6. *If $\mathfrak{N}$ is a subring of $\mathfrak{R}$ and $z \in \mathfrak{N}$, then for any positive integer h either $z^{h-1}\mathfrak{N} > z^h\mathfrak{N}$ or $z^h = 0$.*

Evidently $z^{h-1}\mathfrak{N} \geq z^h\mathfrak{N}$. Suppose that $z^{h-1}\mathfrak{N} = z^h\mathfrak{N}$. Then $z^h = z^h y$ for some y in $\mathfrak{N}$. Let y' be a quasi-inverse of $-y$. Then

$$0 = z^h - z^h y + z^h y' - z^h y y' = z^h + z^h(-y + y' - yy') = z^h.$$

This theorem implies that the radical contains no idempotent element $\neq 0$. It is known that if $\mathfrak{A}$ is a ring with an identity whose lattice of right ideals is completely reducible, then every right ideal $\mathfrak{I}$ in $\mathfrak{A}$ has the form $e\mathfrak{A}$ where e is an idempotent element in $\mathfrak{I}$.[10] Hence we have the following

THEOREM 7. *If $\mathfrak{A}$ is a ring with an identity whose lattice of right (left) ideals is completely reducible, then $\mathfrak{A}$ is semi-simple.*

We shall show next that any ring which is regular[11] in the sense of von Neumann is semi-simple. We recall the definition: $\mathfrak{A}$ is regular if every element a of $\mathfrak{A}$ has a *relative inverse* u such that $aua = a$. Suppose that $a \in \mathfrak{R}$. Then $-ua$ has a quasi-inverse v such that $-ua + v - uav = 0$. Hence

$$0 = -aua + av - auav = -a + av - av = -a.$$

THEOREM 8. *Any regular ring is semi-simple.*

3. The radical of a ring satisfying the descending chain condition.

We suppose now that $\mathfrak{A}$ is a ring for which the descending chain condition for right (left) ideals holds. Let $\mathfrak{N}$ be a two-sided ideal contained in $\mathfrak{R}$ and suppose that $\mathfrak{N}^2 = \mathfrak{N}$. Then if $\mathfrak{N} \neq 0$ there exists a minimum right ideal $\mathfrak{I}$ of $\mathfrak{A}$ with the properties 1) $\mathfrak{I} \leq \mathfrak{N}$, 2) $\mathfrak{I}\mathfrak{N} \neq 0$.

---

[10] See, for example, Jacobson [1], p. 65.

[11] Von Neumann [1].

Let b be an element of $\mathfrak{J}$ such that $b\mathfrak{N} \neq 0$. Then $(b\mathfrak{N})\mathfrak{N} = b\mathfrak{N} \neq 0$ and since $b\mathfrak{N} \leq \mathfrak{J}$ we have $b\mathfrak{N} = \mathfrak{J}$ by the minimality of $\mathfrak{J}$. Since $b \epsilon \mathfrak{J}$ there is an element y in $\mathfrak{N}$ such that $by = b$. As in the proof of Theorem 6 this leads to $b = 0$ contrary to $b\mathfrak{N} \neq 0$. Thus $\mathfrak{N} = 0$. Now the positive integral powers $\mathfrak{R}^k$ of $\mathfrak{R}$ are two-sided ideals and $\mathfrak{R} \geq \mathfrak{R}^2 \geq \cdots$. Hence there is an integer ρ such that $\mathfrak{R}^\rho = \mathfrak{R}^{\rho+1}$. Then for $\mathfrak{N} = \mathfrak{R}^\rho$ we have $\mathfrak{N}^2 = \mathfrak{N}$. Hence $\mathfrak{N} = \mathfrak{R}^\rho = 0$. This proves

THEOREM 9. *If $\mathfrak{A}$ is a ring that satisfies the descending chain condition for right (left) ideals, then the radical of $\mathfrak{A}$ is nilpotent.*

Since any nil ideal is contained in the radical this proves that any nil ideal in a ring that satisfies the descending chain condition for right (left) ideals is nilpotent. It is clear also that $\mathfrak{R}$ coincides with the usual radical defined as the join of all nilpotent ideals.

4. Finitely generated ideals contained in $\mathfrak{R}$.

THEOREM 10. *If $\mathfrak{N}$ is a right ideal with a finite basis contained in the radical $\mathfrak{R}$ then either $\mathfrak{N}\mathfrak{R} < \mathfrak{N}$ or $\mathfrak{N} = 0$.*

Since $\mathfrak{N}$ is a right ideal either $\mathfrak{N}\mathfrak{R} < \mathfrak{N}$ or $\mathfrak{N}\mathfrak{R} = \mathfrak{N}$. Assume that $\mathfrak{N}\mathfrak{R} = \mathfrak{N}$. Let $y_1, \cdots, y_n$ be a basis for $\mathfrak{N}$. Then every element of $\mathfrak{N}$ has the form $\Sigma y_i a_i + \Sigma y_i j_i$ where the $a_i \epsilon \mathfrak{A}$ and the j_i are integers. Since $\mathfrak{N}\mathfrak{R} = \mathfrak{N}$ every element of $\mathfrak{N}$ also has the form $\Sigma y_i z_i$, z_i in $\mathfrak{R}$. In particular $y_1 = \Sigma y_i z_i$. Let z'_1 be the quasi-inverse of $-z_1$. Then we have

$$\begin{aligned} y_1 &= y_1 + y_1(-z_1 + z'_1 - z_1 z'_1) \\ &= (y_1 - y_1 z_1) + (y_1 - y_1 z_1) z'_1 = \sum_2^n y_j (z_j + z_j z'_1). \end{aligned}$$

Hence y_1 can be eliminated from the basis. Similarly every y_i can be eliminated and so $\mathfrak{N} = 0$.

If $\mathfrak{A}$ is an arbitrary ring we define the transfinite powers $\mathfrak{A}^\alpha$ by the conditions 1) $\mathfrak{A}^1 = \mathfrak{A}$, 2) $\mathfrak{A}^{\alpha+1} = \mathfrak{A}^\alpha \mathfrak{A}$, 3) if α is the limit ordinal $\mathfrak{A}^\alpha$ is the join of all $\mathfrak{A}^\beta$ with $\beta < \alpha$. There exists a least ordinal ρ such that $\mathfrak{A}^\rho = \mathfrak{A}^{\rho+1}$. We shall call ρ the *index* of $\mathfrak{A}$ and we shall say that $\mathfrak{A}$ is *transfinite nilpotent* if $\mathfrak{A}^\rho = 0$. Now suppose that $\mathfrak{A}$ is a ring that satisfies the ascending chain condition for right ideals. We recall that this condition is equivalent to the requirement that every ideal has a finite basis. Let $\mathfrak{N} = \mathfrak{R}^\rho$ where ρ is the index of the radical $\mathfrak{R}$ of $\mathfrak{A}$. Then $\mathfrak{N}\mathfrak{R} = \mathfrak{N}$ and so, by Theorem 10, $\mathfrak{N} = \mathfrak{R}^\rho = 0$.

THEOREM 11. *The radical of a ring that satisfies the ascending chain condition for right ideals is transfinite nilpotent.*

5. The radical of a matrix ring. If $\mathfrak{A}$ is an arbitrary ring we denote as usual the ring of $n \times n$ matrices with elements in $\mathfrak{A}$ by $\mathfrak{A}_n$. If $\mathfrak{B}$ is a subring (ideal) in $\mathfrak{A}$ then $\mathfrak{B}_n$ is a subring (ideal) in $\mathfrak{A}_n$. Let $\mathfrak{R}$ be the radical of $\mathfrak{A}$. Then we wish to show that $\mathfrak{R}_n$ is the radical $\mathfrak{R}(\mathfrak{A}_n)$ of $\mathfrak{A}_n$.

LEMMA 2. *Any matrix* $z = (z_{ij})$ *of* $\mathfrak{A}_n$ *in which* z_{11} *is right quasi-regular and the* $z_{ij} = 0$ *for* $i > 1$ *is right quasi-regular in* $\mathfrak{A}_n$.

Since z_{11} is right quasi-regular we know that the ideal $\{x + z_{11}x\} = \mathfrak{A}$. Hence there exist elements z'_{1i} such that $z'_{1i} + z_{11}z'_{1i} = -z_{1i}$. Then if we set $z'_{ij} = 0$ for $i > 1$ we may verify directly that $z' = (z'_{ij})$ is a right quasi-inverse of z.

Consider now the totality $\mathfrak{J}_1$ of matrices with first row consisting of elements in $\mathfrak{R}$ and other rows 0. Then $\mathfrak{J}_1$ is a right ideal. Hence by the lemma $\mathfrak{J}_1 \leq \mathfrak{R}(\mathfrak{A}_n)$. Similarly the totality $\mathfrak{J}_j$ of matrices with j-th row consisting of elements of $\mathfrak{R}$ and other rows 0 is contained in $\mathfrak{R}(\mathfrak{A}_n)$. Since

$$\mathfrak{R}_n = \mathfrak{J}_1 + \cdots + \mathfrak{J}_n, \qquad \mathfrak{R}_n \leq \mathfrak{R}(\mathfrak{A}_n).$$

Conversely let $Y = (y_{ij}) \in \mathfrak{R}(\mathfrak{A}_n)$. If a is any element of $\mathfrak{A}$ we denote the matrix that has a in the (i, j) position and 0's elsewhere by A_{ij}. Let a and b be arbitrary and form the matrix $D = \sum_k A_{kp}YB_{qk}$. Then D is the diagonal matrix $\{ay_{pq}b, \cdots, ay_{pq}b\}$ and $D \in \mathfrak{R}(\mathfrak{A}_n)$. If $D' = (d'_{ij})$ is a right quasi-inverse of D it is easy to see that $d' = d'_{11}$ is a right quasi-inverse of $d = ay_{pq}b$. Evidently this implies that for arbitrary c in $\mathfrak{A}$, and arbitrary integral i, $dc + di$ is right quasi-regular. Hence $d \in \mathfrak{R}$. Since a and b are arbitrary the Corollary to Theorem 5 shows that $y_{pq} \in \mathfrak{R}$ and $Y \in \mathfrak{R}_n$.

THEOREM 12. *If* $\mathfrak{A}$ *is an arbitrary ring and* $\mathfrak{A}_n$ *is the ring of* $n \times n$ *matrices with elements in* $\mathfrak{A}$, *then the radical* $\mathfrak{R}(\mathfrak{R}_n) = \mathfrak{R}_n$, $\mathfrak{R}$ *the radical of* $\mathfrak{A}$.

6. The radical of an algebra. Let $\mathfrak{A}$ be an algebra of possibly infinite order over a field Φ. By an ideal in the algebra $\mathfrak{A}$ we mean, of course, an ideal of the ring $\mathfrak{A}$ that is invariant under the scalar multiplications $x \rightarrow x\alpha$, α in Φ. If $\mathfrak{A}$ has an identity, $x\alpha = x(1\alpha) = (1\alpha)x$ and so any ideal of the ring $\mathfrak{A}$ is an ideal of the algebra $\mathfrak{A}$.

The above discussion is valid without change for an arbitrary algebra $\mathfrak{A}$. Thus the radical $\mathfrak{R}$ of $\mathfrak{A}$ can be defined to be the join of all the quasi-regular right ideals of $\mathfrak{A}$. $\mathfrak{R}$ is also the maximum quasi-regular left ideal of $\mathfrak{A}$. All of our theorems hold for algebras.

An element a of an algebra $\mathfrak{A}$ is *algebraic* if it satisfies a non-trivial algebraic equation with coefficients in the underlying field Φ. An equivalent condition is that a generates a subalgebra A with a finite basis. As in the special case of a field, an element which is not algebraic will be called *transcendental.* If every element of $\mathfrak{A}$ is algebraic, then $\mathfrak{A}$ is algebraic. If a is an algebraic element and A is the subalgebra generated by a then there exists a positive integer h such that $a^{h-1}A = a^hA$. Hence if a is in the radical, by Theorem 6, a is nilpotent. This proves

THEOREM 13. *If $\mathfrak{A}$ is an algebra over a field, the elements of the radical $\mathfrak{R}$ of $\mathfrak{A}$ are either nilpotent or transcendental over Φ.*

If $\mathfrak{A}$ is algebraic every z in $\mathfrak{R}$ is nilpotent. Hence, since $\mathfrak{R}$ contains every nil ideal, we have the following

THEOREM 14. *The radical of an algebraic algebra $\mathfrak{A}$ is the join of all nil right (left) ideals of $\mathfrak{A}$.*

If $\mathfrak{A}$ is commutative, an element z generates a nil ideal if and only if it is nilpotent. Hence we have the

COROLLARY. *If $\mathfrak{A}$ is a commutative algebraic algebra, the radical of $\mathfrak{A}$ is the totality of its nilpotent elements.*

7. The radical of a normed ring. We suppose now that $\mathfrak{A}$ is a normed ring, i. e., $\mathfrak{A}$ is an algebra over the field Φ of complex numbers and for each a in $\mathfrak{A}$ there is defined a non-negative real norm $\| a \|$ such that

1. $\| a \| > 0$ if $a \neq 0$ $\quad \| 0 \| = 0$
2. $\| a + b \| \leq \| a \| + \| b \|$
3. $\| a\alpha \| = \| a \| \cdot | \alpha |$ if $\alpha \in \Phi$
4. $\| ab \| \leq \| a \| \cdot \| b \|$
5. $\mathfrak{A}$ has an identity and $\| 1 \| = 1$
6. $\mathfrak{A}$ is complete relative to the metric $D(a, b) = \| a - b \|$

Following Gelfand we call an element of $\mathfrak{A}$ a *generalized nilpotent element* if $\lim \| z^n \|^{1/n} = 0$.[12] For commutative normed rings Gelfand has defined the radical to be the totality of generalized nilpotent elements. We shall show that this set coincides with the radical as defined here and we shall obtain a similar characterization of the radical for non-commutative normed rings. We prove first the following

THEOREM 15. *The radical of a normed ring $\mathfrak{A}$ is the totality of elements z such that $(za)^n \rightarrow 0$ $((az)^n \rightarrow 0)$ for every a in $\mathfrak{A}$.*

---

[12] Gelfand [1], p. 10.

Since $\mathfrak{A}$ has an identity $\mathfrak{R}$ is the totality of elements z such that $1 + za$ has an inverse for every a in $\mathfrak{A}$. Now suppose that z is an element such that $(za)^n \to 0$ for every a. Then for any α in Φ $\| (z\alpha)^n \| < 1$ for n sufficiently large. Thus $\| z^n \| < \beta^n$ where $\beta = 1/|\alpha|$. We choose α so that $|\alpha| > 1$. The series $1 - z + z^2 - \cdots$ is ultimately dominated by the convergent series $1 + \beta + \beta^2 + \cdots$. Hence $1 - z + z^2 - \cdots$ exists in $\mathfrak{A}$ and this element is the inverse of $1 + z$. Similarly we can show that if $z' = za$, then $1 + z'$ has an inverse. Hence $z \in \mathfrak{R}$. Conversely suppose that $z \in \mathfrak{R}$. Then $1 + z\alpha$ has an inverse $(1 + z\alpha)^{-1}$ for every α in Φ. Using the fact that $(1 + z\alpha)^{-1}$ is an analytic function of α we may prove, exactly as Gelfand has done in the commutative case, that $(z\alpha)^n \to 0$.[13] In particular $z^n \to 0$ and since $\mathfrak{R}$ is an ideal $(za)^n \to 0$ for every a.

We shall call an ideal a *generalized nil ideal* if all of its elements are generalized nilpotent elements. If $z \in \mathfrak{R}$, $(z\alpha)^n \to 0$. Then for n sufficiently large $\| (z\alpha)^n \| < 1$ and $\| z^n \| < \beta^n$ for $\beta = 1/|\alpha|$. Hence $0 \leq \| z^n \|^{1/n} < \beta$. Since β is arbitrary $\lim \| z^n \|^{1/n} = 0$. Thus z is a generalized nilpotent element and $\mathfrak{R}$ is a generalized nil ideal. Next let $\mathfrak{J}$ be an arbitrary generalized nil right ideal. Then if $y \in \mathfrak{J}$, $\| y^n \|^{1/n} \to 0$. Hence $y^n \to 0$. Since $\mathfrak{J}$ is a right ideal, $y' = ya \in \mathfrak{J}$ and $(y')^n \to 0$. By Theorem 15, $y \in \mathfrak{R}$ and so $\mathfrak{J} \leq \mathfrak{R}$. This proves

THEOREM 16. *The radical of a normed ring is a generalized nil ideal that contains every generalized nil right (left) ideal of the ring.*

Let $\mathfrak{A}$ be commutative. Then if z and $a \in \mathfrak{A}$,

$$\| (za)^n \| = \| z^n a^n \| \leq \| z^n \| \cdot \| a^n \| \leq \| z^n \| \cdot \| a \|^n.$$

Hence $\| (za)^n \|^{1/n} \leq \| z^n \|^{1/n} \cdot \| a \|$ and if z is a generalized nilpotent element then za is a generalized nilpotent element. Thus any generalized nilpotent element generates a generalized nil ideal and $\mathfrak{R}$ is the totality of generalized nilpotent elements.

8. Quotient ideals.

We return to the consideration of an arbitrary ring $\mathfrak{A}$. The results that we obtain are also valid for algebras but we shall not state them explicitly for these.

Let $\mathfrak{J}$ be a right ideal in $\mathfrak{A}$. Then if $a \in \mathfrak{A}$ the right multiplication $x \to xa$ determined by a induces an endomorphism $\bar{a}$ in the difference group $\mathfrak{M} = \mathfrak{A} - \mathfrak{J}$. The mapping $\bar{a}$ sends the coset $x + \mathfrak{J}$ into $xa + \mathfrak{J}$. The totality of elements $\bar{a}$ is a subring $\bar{\mathfrak{A}}$ of the ring of endomorphisms of $\mathfrak{M}$ and the correspondence $a \rightarrow \bar{a}$ is a homomorphism between $\mathfrak{A}$ and $\bar{\mathfrak{A}}$. The kernel of this homomorphism is a two-sided ideal $\mathfrak{J}:\mathfrak{A}$ which we shall call the *quotient* of $\mathfrak{J}$ relative to $\mathfrak{A}$. Evidently $\mathfrak{A}(\mathfrak{J}:\mathfrak{A}) \leq \mathfrak{J}$ and if $\mathfrak{A}$ has an identity, $(\mathfrak{J}:\mathfrak{A})$ is the largest two-sided ideal of $\mathfrak{A}$ contained in $\mathfrak{J}$. By the fundamental theorem on homomorphisms $\bar{\mathfrak{A}} \cong \mathfrak{A} - (\mathfrak{J}:\mathfrak{A})$.

---

[13] Gelfand [1], p. 10.

Let $\mathfrak{J}$ be maximal, i. e., $\mathfrak{A} > \mathfrak{J}$ and there is no right ideal $\mathfrak{J}'$ such that $\mathfrak{A} > \mathfrak{J}' > \mathfrak{J}$. Then $\mathfrak{M} = \mathfrak{A} - \mathfrak{J}$ is irreducible relative to $\bar{\mathfrak{A}}$. As usual we call a ring of endomorphisms $\bar{\mathfrak{A}}$ irreducible if the group $\mathfrak{M}$ in which $\bar{\mathfrak{A}}$ acts is irreducible. Let $\bar{\mathfrak{A}} \neq 0$ have this property. Then the totality $\mathfrak{Z}$ of elements z in $\mathfrak{M}$ such that $z\bar{\mathfrak{A}} = 0$ is a subgroup of $\mathfrak{M}$ invariant under $\bar{\mathfrak{A}}$. Hence either $\mathfrak{Z} = 0$ or $\mathfrak{Z} = \mathfrak{M}$. Since $\bar{\mathfrak{A}} \neq 0$, $\mathfrak{Z} \neq \mathfrak{M}$ and so $\mathfrak{Z} = 0$. It follows that if x is any element $\neq 0$ of $\mathfrak{M}$ then $x\bar{\mathfrak{A}} \neq 0$. Since $x\bar{\mathfrak{A}}$ is a subgroup invariant under $\bar{\mathfrak{A}}$, $x\bar{\mathfrak{A}} = \mathfrak{M}$. We use this to prove

THEOREM 17. *Any irreducible ring of endomorphisms is semi-simple.*

Let $\bar{z}$ be an element of the radical of $\bar{\mathfrak{A}}$ and let $x \neq 0$ be arbitrary in $\mathfrak{M}$. If $x\bar{z} \neq 0$, $(x\bar{z})\bar{\mathfrak{A}} = \mathfrak{M}$. Hence there is an $\bar{a}$ in $\bar{\mathfrak{A}}$ such that $x\bar{z}\bar{a} = x$. The element $-\bar{z}\bar{a}$ has a quasi-inverse $\bar{z}'$. Hence

$$x = x - (x\bar{z}\bar{a} - x\bar{z}' + x\bar{z}\bar{a}\bar{z}') = x - x\bar{z}\bar{a} + (x - x\bar{z}\bar{a})\bar{z}' = 0.$$

This contradiction shows that $x\bar{z} = 0$ for all x. Thus $\bar{z} = 0$ and $\bar{\mathfrak{A}}$ is semi-simple.

This theorem has the following important consequence.

COROLLARY. *If $\mathfrak{J}$ is a maximal right ideal $(\mathfrak{J}:\mathfrak{A})$ contains the radical $\mathfrak{R}$ of $\mathfrak{A}$.*

If $(\mathfrak{J}:\mathfrak{A}) = \mathfrak{A}$ there is nothing to prove. Hence suppose that $(\mathfrak{J}:\mathfrak{A}) \neq \mathfrak{A}$. Then $\bar{\mathfrak{A}} \cong \mathfrak{A} - (\mathfrak{J}:\mathfrak{A})$ is an irreducible ring of endomorphisms $\neq 0$. If $z \in \mathfrak{R}$ the coset $\bar{z} = z + (\mathfrak{J}:\mathfrak{A})$ is in the radical of $\mathfrak{A} - (\mathfrak{J}:\mathfrak{A})$. Since $\bar{\mathfrak{A}}$ is semi-simple, $\bar{z} = 0$. Hence $z \in (\mathfrak{J}:\mathfrak{A})$ and $\mathfrak{R} \leq (\mathfrak{J}:\mathfrak{A})$.

9. The radical as intersection of maximal right ideals.

Let $\mathfrak{A}$ be any ring that is not a radical ring. Then $\mathfrak{A}$ contains an element a which is not right quasi-regular. Hence the right ideal $\{x + ax\}$ does not contain a. Moreover, if $\mathfrak{J}$ is a right ideal that contains a and contains the ideal $\{x + ax\}$ then $\mathfrak{J} = \mathfrak{A}$. By using Zorn's maximum principle we may prove

LEMMA 3. *If a is an element of $\mathfrak{A}$ that is not right quasi-regular the right ideal $\{x + ax\}$ can be imbedded in a maximal right ideal.*

This shows, of course, that any ring that is not a radical ring contains maximal right ideals. We assume now that $\mathfrak{A}$ is any ring that contains maximal right ideals. Consider the intersection $\Pi\mathfrak{I}$ of the maximal right ideals $\mathfrak{I}$ of $\mathfrak{A}$. Let $y \in \Pi\mathfrak{I}$. Then y is right quasi-regular. For otherwise $\{x + yx\}$ can be imbedded in the maximal right ideal $\mathfrak{I}$. Then $y \in \mathfrak{I}$ and $\mathfrak{I} = \mathfrak{A}$ contrary to the maximality of $\mathfrak{I}$. Thus every element of $\Pi\mathfrak{I}$ is quasi-regular and since $\Pi\mathfrak{I}$ is a right ideal, $\Pi\mathfrak{I} \leqq \mathfrak{R}$. On the other hand, by the corollary to Theorem 17, $\mathfrak{R} \leq (\mathfrak{I}:\mathfrak{A})$. Hence $\mathfrak{A}\mathfrak{R} \leq \mathfrak{A}(\mathfrak{I}:\mathfrak{A}) \leq \mathfrak{I}$ and $\mathfrak{A}\mathfrak{R} \leq \Pi\mathfrak{I}$. This proves

THEOREM 18. *If $\mathfrak{A}$ is a ring that contains maximal right ideals and $\Pi\mathfrak{I}$ is the intersection of these maximal right ideals, then $\Pi\mathfrak{I} \leq \mathfrak{R}$ and $\mathfrak{A}\mathfrak{R} \leq \Pi\mathfrak{I}$.*

COROLLARY 1. *If $\mathfrak{A}$ is not a radical ring and $\Pi\mathfrak{I}$ is the intersection of the maximal right ideals of $\mathfrak{A}$ then $\Pi\mathfrak{I} \leq \mathfrak{R}$ and $\mathfrak{A}\mathfrak{R} \leq \Pi\mathfrak{I}$.*

COROLLARY 2. *If $\mathfrak{A}$ is a ring with an identity the radical of $\mathfrak{A}$ is the intersection of the maximal right ideals of $\mathfrak{A}$.*

For $\mathfrak{A}\mathfrak{R} = \mathfrak{R}$. Hence $\mathfrak{R} \leq \Pi\mathfrak{I}$ as well as $\Pi\mathfrak{I} \leq \mathfrak{R}$.

The following results also are consequences of Theorem 18:

If $\mathfrak{A}$ is a ring that contains maximal right ideals then $\Pi\mathfrak{I}$ is a two-sided ideal.

For $\mathfrak{A}(\Pi\mathfrak{I}) \leq \mathfrak{A}\mathfrak{R} \leq \Pi\mathfrak{I}$. Hence $\Pi\mathfrak{I}$ is a left ideal as well as a right ideal.

The radical of a normed ring is a closed ideal.

It is known that any maximal ideal $\mathfrak{I}$ is closed.[14] Hence $\mathfrak{R} = \Pi\mathfrak{I}$ is closed.

If $\mathfrak{A}$ is a semi-simple ring, $\Pi\mathfrak{I} = 0$. Suppose, in addition, that $\mathfrak{A}$ satisfies the descending chain condition for right ideals. Then we can find a finite number of maximal right ideals $\mathfrak{I}_j$, $j = 1, \cdots, n$ such that $\Pi\mathfrak{I}_j = 0$. We may suppose that the set $\{\mathfrak{I}_j\}$ is minimal in the sense that $\prod_{k \neq i} \mathfrak{I}_k = \mathfrak{I}'_i \neq 0$ for every i. Then $\mathfrak{I}_i \wedge \mathfrak{I}'_i = 0$ and $\mathfrak{I}'_i \nleqq \mathfrak{I}_i$. Since $\mathfrak{I}_i$ is maximal it follows that $\mathfrak{A} = \mathfrak{I}_i + \mathfrak{I}'_i$. Hence $\mathfrak{I}'_i$ is isomorphic to the difference $\mathfrak{A}$-group $\mathfrak{A} - \mathfrak{I}_i$ and $\mathfrak{I}'_i$ is minimal. Using a simple lattice-theoretic argument we can conclude that $\mathfrak{A} = \mathfrak{I}'_1 \oplus \cdots \oplus \mathfrak{I}'_n$.[15] This proves the well-known

THEOREM 19. *If $\mathfrak{A}$ is a semi-simple ring that satisfies the descending chain condition for right ideals, $\mathfrak{A}$ is a direct sum of a finite number of minimal right ideals.*

---

[14] Gelfand [1], p. 8.

[15] Jacobson [1], p. 35.

Suppose again that $\mathfrak{A}$ is any ring that contains maximal right ideals. Then, by the corollary to Theorem 17, $\mathfrak{R} \leqq \Pi(\mathfrak{I}:\mathfrak{A})$ for all maximal $\mathfrak{I}$. Conversely let $y \in \Pi(\mathfrak{I}:\mathfrak{A})$. Then $\mathfrak{A}y \leq \Pi\mathfrak{I} \leq \mathfrak{R}$. Hence $\mathfrak{A}y\mathfrak{A} \leq \mathfrak{R}$ and this implies that $y \in \mathfrak{R}$. We have therefore proved

THEOREM 20. *Let $\mathfrak{A}$ be an arbitrary ring that contains maximal right ideals. Then the radical of $\mathfrak{A}$ is the intersection $\Pi(\mathfrak{I}:\mathfrak{A})$ where $\mathfrak{I}$ ranges over the maximal right ideals of $\mathfrak{A}$.*

COROLLARY. *If $\mathfrak{A}$ is not a radical ring, $\mathfrak{R} = \Pi(\mathfrak{I}:\mathfrak{A})$ where $\mathfrak{I}$ ranges over the maximal right ideals.*

The results of this section hold also for left ideals. An interesting consequence of the second corollary to Theorem 18 is that if $\mathfrak{A}$ is a ring with an identity then the intersection of the maximal right ideals of $\mathfrak{A}$ coincides with the intersection of the maximal left ideals of $\mathfrak{A}$.

10. Primitive rings.

Definition 3. A ring $\mathfrak{A}$ is *primitive* if $\mathfrak{A}$ contains a maximal right ideal $\mathfrak{I}$ whose quotient $(\mathfrak{I}:\mathfrak{A}) = 0$.

Let $\mathfrak{A}$ be of this type and as before let $\bar{\mathfrak{A}}$ denote the ring of endomorphisms $x + \mathfrak{I} \to xa + \mathfrak{I}$ in $\mathfrak{M} = \mathfrak{A} - \mathfrak{I}$. Then $\bar{\mathfrak{A}}$ is irreducible and $\bar{\mathfrak{A}} \cong \mathfrak{A} - (\mathfrak{I}:\mathfrak{A}) = \mathfrak{A}$. Thus any primitive ring is isomorphic to an irreducible ring of endomorphisms.

Conversely suppose that $\bar{\mathfrak{A}} \neq 0$ is an irreducible ring of endomorphisms acting in $\mathfrak{M}$. Let x be an element $\neq 0$ in $\mathfrak{M}$ and let $\mathfrak{I}_x$ denote the totality of elements b of $\bar{\mathfrak{A}}$ such that $xb = 0$. Evidently $\mathfrak{I}_x$ is a right ideal in $\bar{\mathfrak{A}}$. We have seen that $x\bar{\mathfrak{A}} = \mathfrak{M}$. Hence $\mathfrak{I}_x < \bar{\mathfrak{A}}$. Let $\bar{a}$ be an element of $\bar{\mathfrak{A}}$ not in $\mathfrak{I}_x$ and let $\bar{c}$ be arbitrary. Then $x\bar{a} \neq 0$. Hence $(x\bar{a})\bar{\mathfrak{A}} = \mathfrak{M}$ and so there exists an element $\bar{u}$ such that $x\bar{a}\bar{u} = x\bar{c}$. Thus $\bar{c} - \bar{a}\bar{u} \in \mathfrak{I}_x$ and $\bar{\mathfrak{A}} = \mathfrak{I}_x + \bar{a}\bar{\mathfrak{A}}$ for any $\bar{a}$ not in $\mathfrak{I}_x$. This proves that $\mathfrak{I}_x$ is maximal. If $\bar{c} \neq 0 \in \bar{\mathfrak{A}}$ there is a y in $\mathfrak{M}$ such that $y\bar{c} \neq 0$. There exists an $\bar{a}$ such that $x\bar{a} = y$. It follows that $x\bar{a}\bar{c} \neq 0$. Thus for any $\bar{c} \neq 0$ there exists an $\bar{a}$ such that $\bar{a}\bar{c} \notin \mathfrak{I}_x$. On the other hand, if $\bar{c} \in (\mathfrak{I}_x:\bar{\mathfrak{A}})$, $\bar{a}\bar{c} \in \mathfrak{I}_x$ for all $\bar{a}$. Hence $\bar{c} = 0$ and $(\mathfrak{I}_x:\bar{\mathfrak{A}}) = 0$.

THEOREM 21. *A necessary and sufficient condition that $\mathfrak{A}$ be primitive is that $\mathfrak{A}$ be isomorphic to an irreducible ring of endomorphisms.*

If $\bar{\mathfrak{B}}$ is a two-sided ideal $\neq 0$ in the irreducible ring $\bar{\mathfrak{A}}$, the argument that we have used before for $\bar{\mathfrak{A}}$ shows that $x\bar{\mathfrak{B}} = \mathfrak{M}$ for any $x \neq 0$. Thus $\bar{\mathfrak{B}}$ is irreducible. Hence we have the following

THEOREM 22. *Any two-sided ideal in a primitive ring is primitive.*

For a deeper study of primitive rings we require a theorem on irreducible rings of endomorphisms recently proved by the author. We recall first that if $\bar{\mathfrak{A}}$ is irreducible in $\mathfrak{M}$ then the totality of endomorphisms of $\mathfrak{M}$ that commute with all the $\bar{a} \in \bar{\mathfrak{A}}$ is a division subring $\mathfrak{D}$ of the ring of endomorphisms of $\mathfrak{M}$.[16] Since $\mathfrak{D}$ contains the identity mapping, we may regard $\mathfrak{M}$ as a vector space over $\mathfrak{D}$.

We call a set $\bar{\mathfrak{A}}$ of linear transformations in a vector space $\mathfrak{M}$ over $\mathfrak{D}$ a *dense set* if for any finite set $x_1, \cdots, x_n$ of $\mathfrak{D}$-independent vectors and any finite set $y_1, \cdots, y_n$ of arbitrary vectors there exists a linear transformation $\bar{a} \in \bar{\mathfrak{A}}$ such that $x_i\bar{a} = y_i$. An equivalent condition is that if $\mathfrak{S}$ is any finite dimensional subspace of $\mathfrak{M}$ and A is any linear transformation in $\mathfrak{S}$ over $\mathfrak{D}$, then A can be extended to a linear transformation $\bar{a} \in \bar{\mathfrak{A}}$ of $\mathfrak{M}$ over $\mathfrak{D}$.[17] Hence if $\mathfrak{M}$ is finite dimensional $\bar{\mathfrak{A}} = \mathfrak{L}$ the complete ring of linear transformations in $\mathfrak{M}$ over $\mathfrak{D}$. We have proved elsewhere the following

THEOREM 23. *If $\bar{\mathfrak{A}} \neq 0$ is an irreducible ring of endomorphisms in the commutative group $\mathfrak{M}$ and $\mathfrak{D}$ is the division ring of endomorphisms commutative with the $\bar{a} \in \bar{\mathfrak{A}}$, then $\bar{\mathfrak{A}}$ is a dense ring of linear transformations in the vector space $\mathfrak{M}$ over $\mathfrak{D}$.*

The converse is trivial. Thus the concepts of primitive ring, irreducible ring of endomorphisms and dense ring of linear transformations may be used interchangeably.

If $\bar{\mathfrak{A}}$ is a dense ring of linear transformations in $\mathfrak{M}$ over $\mathfrak{D}$ and $x_1, x_2, x_3, \cdots$ are $\mathfrak{D}$-independent, then

$$\mathfrak{J}_{x_1} > (\mathfrak{J}_{x_1} \wedge \mathfrak{J}_{x_2}) > (\mathfrak{J}_{x_1} \wedge \mathfrak{J}_{x_2} \wedge \mathfrak{J}_{x_3}) > \cdots$$

where $\mathfrak{J}_x$ denotes the right ideal of annihilators of x. Hence if $\mathfrak{M}$ is infinite dimensional, then $\bar{\mathfrak{A}}$ does not satisfy the descending chain condition for right ideals. Thus if $\bar{\mathfrak{A}}$ satisfies the descending chain condition $\mathfrak{M}$ is finite dimensional over $\mathfrak{D}$ and therefore $\bar{\mathfrak{A}} = \mathfrak{L}$ the complete ring of linear transformations in $\mathfrak{M}$ over $\mathfrak{D}$. This proves

THEOREM 24. *Any primitive ring that satisfies the descending chain condition for right ideals is isomorphic to the complete ring of linear transformations of a suitable finite dimensional vector space over a division ring.*

---

[16] This is Schur's lemma. See Jacobson [1], p. 57.

[17] For the definitions and results on dense rings of linear transformations quoted here, see Jacobson [2].

Since the complete ring of linear transformations in a finite dimensional vector space is simple we have the

COROLLARY. *Any primitive ring that satisfies the descending chain condition for right ideals is simple.*

11. Structure of semi-simple rings. We have seen that if $\mathfrak{A}$ is not a radical ring, the radical $\mathfrak{R} = \Pi(\mathfrak{I} : \mathfrak{A})$ for the maximal right ideals $\mathfrak{I}$. The difference rings $\mathfrak{A} - (\mathfrak{I} : \mathfrak{A})$ are isomorphic to irreducible rings of endomorphisms. Hence they are primitive. In particular we have the following

THEOREM 25. *If* $\mathfrak{A}$ *is a semi-simple ring* $\mathfrak{A}$ *contains two-sided ideals* $\mathfrak{B}$ *such that* 1) *each* $\mathfrak{A} - \mathfrak{B}$ *is primitive and* 2) $\Pi\mathfrak{B} = 0$.

The converse holds also. For if $\mathfrak{B}$ is a two-sided ideal in a ring and $\mathfrak{A} - \mathfrak{B}$ is semi-simple then $\mathfrak{B} \geq \mathfrak{R}$. Since any primitive ring is semi-simple each ideal $\mathfrak{B}$ such that $\mathfrak{A} - \mathfrak{B}$ is primitive contains $\mathfrak{R}$. Hence if $\Pi\mathfrak{B} = 0$ $\mathfrak{R} = 0$.

Theorem 25 reduces, to a certain extent, the study of semi-simple rings to that of primitive rings and hence to that of dense rings of linear transformations in a vector space over a division ring. This will be clearer in the next section where we derive a certain alternate form of this theorem. Before doing this we obtain some consequences of the theorem in the present form. We note first

THEOREM 26. *Any two-sided ideal in a semi-simple ring is semi-simple.*

Let $\mathfrak{A}_1$ be a two-sided ideal in the semi-simple ring $\mathfrak{A}$ and let $\{\mathfrak{B}\}$ be a set of two-sided ideals in $\mathfrak{A}$ such that 1) each $\mathfrak{A} - \mathfrak{B}$ is primitive and 2) $\Pi\mathfrak{B} = 0$. Let $\mathfrak{B}_1 = \mathfrak{A}_1 \wedge \mathfrak{B}$. Then $\mathfrak{A}_1 - (\mathfrak{A}_1 \wedge \mathfrak{B}) \cong (\mathfrak{A}_1 + \mathfrak{B}) - \mathfrak{B}$ a two-sided ideal in the primitive ring $\mathfrak{A} - \mathfrak{B}$. It follows that $\mathfrak{A}_1 - \mathfrak{B}_1$ is primitive. Evidently $\Pi\mathfrak{B} = 0$. Hence $\mathfrak{A}_1$ is semi-simple.

Next let $\mathfrak{A}$ be a semi-simple ring that satisfies the descending chain condition for right ideals. Let $\{\mathfrak{B}\}$ be a set of two-sided ideals satisfying the conditions of Theorem 25. We may replace the set $\{\mathfrak{B}\}$ by a finite set $\mathfrak{B}_i$, $i = 1, \cdots, s$, and we may suppose that the set is minimal in the sense that $\mathfrak{B}'_i = \prod_{j \neq i} \mathfrak{B}_j \neq 0$. Since $\mathfrak{A} - \mathfrak{B}$ is primitive and satisfies the descending chain condition for right ideals, $\mathfrak{A} - \mathfrak{B}$ is simple and so $\mathfrak{B}$ is a maximal two-sided ideal in $\mathfrak{A}$. It follows that $\mathfrak{B}'_i + \mathfrak{B}_i = \mathfrak{A}$. Hence $\mathfrak{B}'_i \cong \mathfrak{A} - \mathfrak{B}_i$ is simple. As for one-sided ideals we can prove that $\mathfrak{A} = \mathfrak{B}'_1 \oplus \cdots \oplus \mathfrak{B}'_s$. This proves the Wedderburn-Artin structure theorem:

THEOREM 27. *Any semi-simple ring that satisfies the descending chain condition for right ideals is isomorphic to a direct sum of a finite number of simple rings.*

By the preceding section the structure of the component ring $\mathfrak{B}'_i$ is that of a ring of linear transformations in a finite dimensional vector space over a division ring. This is the second fundamental Wedderburn-Artin structure theorem.

12. Infinite direct sums. Let S be an arbitrary set. We call the elements of S points and we suppose that to each point P there is associated a ring $\mathfrak{E}_P$. Now let $\mathfrak{S} = (S, \mathfrak{E}_P)$ be the totality of functions $f(P)$ with domain S and with $f(P)$ in $\mathfrak{E}_P$. If $f(P)$ and $g(P) \epsilon \mathfrak{S}$ we define $f + g$ by $(f + g)(P) = f(P) + g(P)$ and we define $fg(P) = f(P)g(P)$. Then it is readily verified that $\mathfrak{S}$ is a ring. We shall call it the *complete direct sum* of the rings $\mathfrak{E}_P$. Let $\mathfrak{S}_0$ be the subset of functions $f(P)$ such that $f(P) = 0$ for all but a finite number of the points P. Then $\mathfrak{S}_0$ is a subring of $\mathfrak{S}$. We call it the (discrete) *direct sum* of the rings $\mathfrak{E}_P$. Let $\overline{\mathfrak{E}}_Q$ be the subset of $\mathfrak{S}_0$ of elements $f(P)$ such that $f(P) = 0$ for all $P \neq Q$. Then 1) $\overline{\mathfrak{E}}_Q$ is a two-sided ideal in $\mathfrak{S}_0$, 2) $\mathfrak{S}_0 = \Sigma\overline{\mathfrak{E}}_Q$ where $\Sigma\overline{\mathfrak{E}}_Q$ denotes the totality of finite sums of the elements of the ideals $\overline{\mathfrak{E}}_Q$ and 3) $\overline{\mathfrak{E}}_Q \wedge \sum_{P\neq Q} \overline{\mathfrak{E}}_P = 0$. Conversely these conditions are sufficient that a ring be isomorphic to a direct sum of rings isomorphic to the ideals $\overline{\mathfrak{E}}_Q$.

If $\mathfrak{A}$ is a subring of the complete direct sum $\mathfrak{S}$ we define the *component* $\mathfrak{A}_P$ *of* $\mathfrak{A}$ *at* P to be the totality of elements $f(P)$ of $\mathfrak{E}_P$ for f in $\mathfrak{A}$. It is clear that $\mathfrak{A}_P$ is a subring of $\mathfrak{E}_P$ and the correspondence $f \rightarrow f(P)$ is a homomorphism between $\mathfrak{A}$ and $\mathfrak{A}_P$. The kernel of this homomorphism is the two-sided ideal $\mathfrak{B}_P$ of elements g such that $g(P) = 0$. Hence $\Pi\mathfrak{B}_P = 0$. Moreover, $\mathfrak{A}_P \cong \mathfrak{A} - \mathfrak{B}_P$.

We suppose, conversely, that S is a set of two-sided ideals $P = \mathfrak{B}$ such that $\Pi\mathfrak{B} = 0$. Let $\mathfrak{A}_P = \mathfrak{A} - \mathfrak{B}$ and form the complete direct sum $\mathfrak{S} = (S, \mathfrak{A}_P)$. Then each a in $\mathfrak{A}$ determines an element $\bar{a}$ in $\mathfrak{S}$ such that $\bar{a}(P)$ is the coset $a + \mathfrak{B}$. The set $\overline{\mathfrak{A}}$ of the $\bar{a}$ is a subring of $\mathfrak{S}$ and the correspondence $a \rightarrow \bar{a}$ is a homomorphism. It is clear that component $\overline{\mathfrak{A}}_P = \mathfrak{A}_P$ and that $\bar{a}(P) = 0$ if and only if $a \epsilon \mathfrak{B}$. Thus $\bar{a} = 0$ if and only if $a \epsilon \Pi\mathfrak{B}$ and since $\Pi\mathfrak{B} = 0$ we see that the correspondence $a \rightarrow \bar{a}$ is, in fact, an isomorphism between $\mathfrak{A}$ and $\overline{\mathfrak{A}}$. Using these remarks we may formulate Theorem 25 and its converse as follows:

THEOREM 28. *A necessary and sufficient condition that a ring* $\mathfrak{A}$ *be semi-*

simple is that $\mathfrak{A}$ *be isomorphic to a subring* $\overline{\mathfrak{A}}$ *of a complete direct sum* $(S, \overline{\mathfrak{A}}_P)$ *where the components* $\overline{\mathfrak{A}}_P$ *of* $\overline{\mathfrak{A}}$ *are primitive.*

This theorem serves as a substitute in the general theory for the Wedderburn-Artin Theorem 27. A number of special cases of the theorem are well-known. We cite, for example, Stone's theorem on the representability of a Boolean ring as a subring of a direct sum of the fields $\mathfrak{I}_2$ of residues mod 2 and Montgomery and McCoy's generalization of this theorem to commutative rings in which every element a satisfies the equations $pa = 0$, $a^p = a$ for a fixed prime p. These results follow from Theorem 28 and the easily proved fact that any primitive commutative ring is a field.

13. Primitive rings that contain minimal ideals. If $\mathfrak{A}$ is an arbitrary ring it is well-known that any minimal right ideal $\mathfrak{I}$ in $\mathfrak{A}$ either has the form $e\mathfrak{A}$ where e is an idempotent element or $\mathfrak{I}^2 = 0$.[18] If $\mathfrak{I}$ is a minimal right ideal and a is arbitrary $a\mathfrak{I}$ is either 0 or a minimal right ideal. It follows that the join $\Sigma\mathfrak{I}$ of all the minimal right ideals is a two-sided ideal.[19]

If $\overline{\mathfrak{A}}$ is an irreducible ring of endomorphisms in the commutative group $\mathfrak{M}$ and $\overline{\mathfrak{B}}$ is a two-sided ideal $\neq 0$ in $\overline{\mathfrak{A}}$ we have seen that $x\overline{\mathfrak{B}} = \mathfrak{M}$ for every $x \neq 0$ in $\mathfrak{M}$. $\overline{\mathfrak{B}}$ is primitive also. If $\overline{\mathfrak{B}}_1$ and $\overline{\mathfrak{B}}_2$ are two-sided ideals $\neq 0$ then the equation $x\overline{\mathfrak{B}}_1 = \mathfrak{M}$ implies that $\overline{\mathfrak{B}}_1\overline{\mathfrak{B}}_2 \neq 0$. We state this as

LEMMA 4. *If* $\mathfrak{B}_1, \mathfrak{B}_2$ *are two-sided ideals* $\neq 0$ *in the primitive ring* $\mathfrak{A}$, $\mathfrak{B}_1\mathfrak{B}_2 \neq 0$.

This, of course, implies that $\mathfrak{B}_1 \wedge \mathfrak{B}_2 \neq 0$.

Suppose now that $\mathfrak{A}$ is a primitive ring that contains minimal right ideals. Since $\mathfrak{A}$ is semi-simple any minimal right ideal $\mathfrak{I}$ of $\mathfrak{A}$ has the form $e\mathfrak{A}$, $e^2 = e \neq 0$. Let $\mathfrak{F}$ be the two-sided ideal $\Sigma\mathfrak{I}$ for all minimal right ideals $\mathfrak{I}$. We assert that $\mathfrak{F}$ is a minimal two-sided ideal. For let $\mathfrak{F}_1$ be a two-sided ideal such that $\mathfrak{F}_1 < \mathfrak{F}$. There exists a minimal right ideal $\mathfrak{I} \nleqq \mathfrak{F}_1$. Since $\mathfrak{I} \wedge \mathfrak{F}_1$ is a right ideal and $\mathfrak{I}$ is minimal it follows that $\mathfrak{I} \wedge \mathfrak{F}_1 = 0$. Hence also $\mathfrak{I}\mathfrak{F}_1 = 0$. Since $\mathfrak{I}$ is not nilpotent $\mathfrak{A}\mathfrak{I} \neq 0$. Evidently $\mathfrak{A}\mathfrak{I}$ is a two-sided ideal and since $(\mathfrak{A}\mathfrak{I})\mathfrak{F}_1 = 0$, $\mathfrak{F}_1 = 0$ by the lemma. The minimality of $\mathfrak{F}$ has therefore been established. It follows from Lemma 4 that $\mathfrak{F}$ is contained in every two-sided ideal $\mathfrak{B} \neq 0$ of $\mathfrak{A}$. We wish to show next that $\mathfrak{F}$ is a simple ring. Let $\mathfrak{F}_1 < \mathfrak{F}$ now denote a two-sided ideal of $\mathfrak{F}$. Then $\mathfrak{F}\mathfrak{F}_1\mathfrak{F} \leqq \mathfrak{F}_1 < \mathfrak{F}$ and $\mathfrak{F}\mathfrak{F}_1\mathfrak{F}$ is a two-sided ideal of $\mathfrak{A}$. Hence $\mathfrak{F}\mathfrak{F}_1\mathfrak{F} = 0$ and since $\mathfrak{F}$ is primitive $\mathfrak{F}_1 = 0$.

---

[18] See, for example, Jacobson [1], p. 64.

[19] This ideal has been called the anti-radical by Baer. See Baer [1], p. 544.

THEOREM 29. *Let $\mathfrak{A}$ be a primitive ring that contains minimal right ideals and let $\mathfrak{F} = \Sigma\mathfrak{J}$ the join of all the minimal right ideals of $\mathfrak{A}$. Then $\mathfrak{F}$ is a two-sided ideal that is contained in every two-sided ideal $\neq 0$ of $\mathfrak{A}$ and $\mathfrak{F}$ is a simple ring.*

If $\mathfrak{B}$ is a two-sided ideal $\neq 0$ of $\mathfrak{A}$, $\mathfrak{B}$ contains every minimal right ideal $\mathfrak{J}$ of $\mathfrak{A}$. It is easy to see by the above reasoning that $\mathfrak{J}$ is minimal in $\mathfrak{B}$. Hence $\mathfrak{B}$ satisfies the same conditions as $\mathfrak{A}$.

Again let $\bar{\mathfrak{A}}$ be an irreducible ring of endomorphisms in a commutative group $\mathfrak{M}$ and let $\mathfrak{D}$ be the division ring of endomorphisms in $\mathfrak{M}$ commutative with the elements of $\bar{\mathfrak{A}}$. We call a linear transformation A in a vector space $\mathfrak{M}$ over $\mathfrak{D}$ *finite valued* if the image space $\mathfrak{M}A$ is finite dimensional. We have shown elsewhere that if $\bar{\mathfrak{A}}$ contains finite valued linear transformations $\neq 0$ then $\bar{\mathfrak{A}}$ contains minimal right ideals.[20] Conversely suppose that $\bar{\mathfrak{A}}$ contains a minimal right ideal $\mathfrak{J}$. Then $\mathfrak{J} = \bar{e}\bar{\mathfrak{A}}$, $\bar{e}^2 = \bar{e} \neq 0$. We wish to show that $\mathfrak{M}\bar{e}$ is one dimensional. For, if not, there exist two $\mathfrak{D}$-independent elements x, y of $\mathfrak{M}$ such that $x\bar{e} = x$, $y\bar{e} = y$. By the density of $\bar{\mathfrak{A}}$ there exists a $\bar{b}$ in $\bar{\mathfrak{A}}$ such that $x\bar{b} = 0$, $y\bar{b} \neq 0$. Then $\bar{c} = \bar{e}\bar{b}$ satisfies $x\bar{c} = 0$, $y\bar{c} \neq 0$ and $\bar{c} \in \mathfrak{J}$. Hence if $\mathfrak{J}_x$ denotes the annihilator of x, $\mathfrak{J}_x \wedge \mathfrak{J} \neq 0$. Since $\mathfrak{J}$ is minimal $\mathfrak{J} = \mathfrak{J} \wedge \mathfrak{J}_x$. But this contradicts the fact that $\bar{e} \in \mathfrak{J}$ and $x\bar{e} \neq 0$.

We have shown also that if $\bar{\mathfrak{A}}$ contains finite valued transformations $\neq 0$ the totality $\bar{\mathfrak{F}}$ of these transformations is a two-sided ideal contained in every two-sided ideal $\neq 0$ of $\bar{\mathfrak{A}}$.[21] Thus $\bar{\mathfrak{F}} = \Sigma\mathfrak{J}$ for all minimal right ideals $\mathfrak{J}$. This completes the proof of the following

THEOREM 30. *Let $\bar{\mathfrak{A}} \neq 0$ be an irreducible ring of endomorphisms in $\mathfrak{M}$ and let $\mathfrak{D}$ be the division ring of endomorphisms commutative with the elements of $\bar{\mathfrak{A}}$. Then $\bar{\mathfrak{A}}$ contains minimal right ideals if and only if it contains finite valued linear transformations $\neq 0$ of $\mathfrak{M}$ over $\mathfrak{D}$. If the condition is satisfied the minimal two-sided ideal $\bar{\mathfrak{F}}$ coincides with the totality of finite valued transformations in $\bar{\mathfrak{A}}$.*

We recall that a homomorphism $a \rightarrow \bar{a}$ between an abstract ring $\mathfrak{A}$ and a ring of endomorphisms $\bar{\mathfrak{A}}$ in a commutative group $\mathfrak{M}$ is called a *representation* of $\mathfrak{A}$. The representation is *irreducible* if $\bar{\mathfrak{A}}$ is irreducible. Two representations are *equivalent* if the groups in which the endomorphisms act are $\mathfrak{A}$-isomorphic. We have seen that the fundamental fact about a primitive ring is that it possesses a $(1 - 1)$ irreducible representation.

---

[20] Jacobson [2], p. 232.

[21] Jacobson [2], p. 230.

Suppose now that $\mathfrak{A}$ is a primitive ring that contains a minimal right ideal $\mathfrak{J}$ and let $a \rightarrow \bar{a}$ be any $(1-1)$ irreducible representation of $\mathfrak{A}$. Let $\bar{\mathfrak{A}}$ be the corresponding ring of endomorphisms and $\mathfrak{M}$ the group in which they act. The image $\bar{\mathfrak{J}}$ of $\mathfrak{J}$ is $\neq 0$. Hence there exists an x in $\mathfrak{M}$ such that $x\bar{\mathfrak{J}} \neq 0$. Because of the irreducibility of $\mathfrak{M}$, $x\bar{\mathfrak{J}} = \mathfrak{M}$. It can be verified directly that the correspondence $u \rightarrow xu$, u in $\mathfrak{J}$, is an $\mathfrak{A}$-isomorphism between $\mathfrak{J}$ and $\mathfrak{M}$. Thus the representation in $\mathfrak{M}$ is equivalent to the representation by the right multiplications in $\mathfrak{J}$. This proves

THEOREM 31. *If $\mathfrak{A}$ is a primitive ring that contains minimal right ideals, any two $(1-1)$ irreducible representations of $\mathfrak{A}$ are equivalent.*

Now let $\bar{\mathfrak{A}}_1$ and $\bar{\mathfrak{A}}_2$ be dense rings of linear transformations in $\mathfrak{M}_1$ over $\mathfrak{D}_1$ and $\mathfrak{M}_2$ over $\mathfrak{D}_2$ respectively. Suppose that the $\bar{\mathfrak{A}}_i$ are isomorphic under a correspondence $\bar{a}_1 \rightarrow \bar{a}_2$ and that the $\bar{\mathfrak{A}}_i$ contain minimal right ideals. Then if $\mathfrak{A}$ is the abstract primitive ring isomorphic to the $\bar{\mathfrak{A}}_i$ we have two irreducible representations $a \rightarrow \bar{a}_1$, $a \rightarrow \bar{a}_2$ of $\mathfrak{A}$. Since these representations are equivalent we have a $(1-1)$ correspondence $S: x_1 \rightarrow x_2 = x_1 S$ between $\mathfrak{M}_1$ and $\mathfrak{M}_2$ such that

$$(x_1 + y_1)S = x_1 S + y_1 S, \qquad (x_1 \bar{a}_1)S = (x_1 S)\bar{a}_2.$$

The latter equation may be written in the form $\bar{a}_2 = S^{-1}\bar{a}_1 S$. Now it is easy to see that $\mathfrak{D}_i$ is the complete set of endomorphisms in $\mathfrak{M}_i$ commutative with $\bar{\mathfrak{A}}_i$.[22] Hence the correspondence $\alpha_1 \rightarrow S^{-1}\alpha_1 S = \alpha_2$ for α_1 in $\mathfrak{D}_1$ is an isomorphism between $\mathfrak{D}_1$ and $\mathfrak{D}_2$. This shows that the two vector spaces are isomorphic. If we identify these spaces, setting $\mathfrak{M}_i = \mathfrak{M}$, $\mathfrak{D}_i = \mathfrak{D}$, then S becomes a semi-linear transformation in $\mathfrak{M}$ over $\mathfrak{D}$.

THEOREM 32. *Two dense rings of linear transformations that contain minimal right ideals can not be isomorphic unless the spaces in which they act are isomorphic. If $\bar{\mathfrak{A}}_1$ and $\bar{\mathfrak{A}}_2$ are two dense rings in $\mathfrak{M}$ over $\mathfrak{D}$ and the $\bar{\mathfrak{A}}_i$ contain minimal right ideals and are isomorphic under a correspondence $\bar{a}_1 \rightarrow \bar{a}_2$ then there exists a $(1-1)$ semi-linear transformation S in $\mathfrak{M}$ over $\mathfrak{D}$ such that $\bar{a}_2 = S^{-1}\bar{a}_1 S$ for all $\bar{a}_1$.*

This result shows that for primitive rings that contain minimal right ideals we have precisely the same uniqueness of representation as a dense ring of linear transformations as we have in the classical case of simple rings that satisfy the descending chain condition.

---

[22] Jacobson [2], p. 233.

14. Atomic semi-simple rings. Let $\mathfrak{A}$ be any ring that contains minimal right ideals and let $\mathfrak{J}$ be the join $\Sigma\mathfrak{I}$ of all the minimal right ideals $\mathfrak{I}$ of $\mathfrak{A}$. The right multiplications $u \to ua$ in $\mathfrak{I}$ form an irreducible ring and therefore a semi-simple ring. It follows that if $\mathfrak{R}$ is the radical of $\mathfrak{A}$ then $\mathfrak{I}\mathfrak{R} = 0$. Hence $\mathfrak{J}\mathfrak{R} = 0$.

We shall call a ring $\mathfrak{A}$ *atomic* if $\mathfrak{A} = \mathfrak{J}$. If $\mathfrak{A}$ is atomic $\mathfrak{A}$ is semi-simple if and only if $\mathfrak{A}$ contains no nilpotent ideals. Let $\mathfrak{A}$ be of this type and let $\mathfrak{I}$ be a minimal right ideal in $\mathfrak{A}$. Then we have the following

LEMMA 5. *If $\mathfrak{A}$ is atomic and semi-simple and $\mathfrak{I}$ is a minimal right ideal in $\mathfrak{A}$ then $\mathfrak{A}\mathfrak{I} \geqq \mathfrak{I}$ and $\mathfrak{A}\mathfrak{I}$ is a simple ring.*

If $\mathfrak{B}$ is any two-sided ideal of $\mathfrak{A}$ either $\mathfrak{B} \geqq \mathfrak{I}$ or $\mathfrak{I} \wedge \mathfrak{B} = 0 = \mathfrak{I}\mathfrak{B}$. Now if $\mathfrak{I}(\mathfrak{A}\mathfrak{I}) = 0$, $\mathfrak{I}^3 = 0$ contrary to the semi-simplicity. Hence $\mathfrak{A}\mathfrak{I} \geqq \mathfrak{I}$. If $\mathfrak{B}$ is any two-sided ideal of $\mathfrak{A}$ properly contained in $\mathfrak{A}\mathfrak{I}$, $\mathfrak{B} \ngeqq \mathfrak{I}$. Hence $\mathfrak{I}\mathfrak{B} = 0$, $(\mathfrak{A}\mathfrak{I})\mathfrak{B} = 0$ and $\mathfrak{B}^2 = 0$. Hence $\mathfrak{B} = 0$. Thus $\mathfrak{A}\mathfrak{I}$ is a minimal two-sided ideal of $\mathfrak{A}$. It follows readily that $\mathfrak{A}\mathfrak{I}$ is a simple ring.

Let $\{\mathfrak{A}_\alpha\}$ be the set of distinct two-sided ideals $\mathfrak{A}\mathfrak{I}$, $\mathfrak{I}$ minimal. Each $\mathfrak{A}_\alpha$ is simple. Hence if $\alpha \neq \beta$, $\mathfrak{A}_\alpha\mathfrak{A}_\beta = 0$. It follows that $\mathfrak{A}_\alpha \wedge \sum_{\beta \neq \alpha} \mathfrak{A}_\beta = 0$. By the lemma $\mathfrak{A} = \Sigma\mathfrak{A}_\alpha$. Hence we have proved

THEOREM 33. *If $\mathfrak{A}$ is atomic and semi-simple $\mathfrak{A}$ is isomorphic to a direct sum of simple rings.*

Each component $\mathfrak{A}\mathfrak{I}$ is a join of minimal right ideals $a\mathfrak{I}$ of $\mathfrak{I}$. It is easy to see that these ideals are also minimal in $\mathfrak{A}\mathfrak{I}$. Hence by Theorem 29 each $\mathfrak{A}\mathfrak{I}$ is isomorphic to a dense ring of finite-valued linear transformations.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY

A. A. Albert:
[1] *Modern Higher Algebra*, Chicago, 1937.

R. Baer:
[1] "Radical ideals," *American Journal of Mathematics*, vol. LXV (1943), pp. 537-568.

G. Birkhoff:

[1] "The radical of a group with operators," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 751-753.

[2] "Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-769.

I. Gelfand:

[1] "Normierte Ringe," *Math. Sbornik*, vol. 9 (51, 1941), pp. 1-23.

N. Jacobson:

[1] "The Theory of Rings," *Mathematical Surveys*, vol. 2 (New York, 1943).

[2] "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.

J. Levitzki:

[1] "On the radical of a general ring," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 461-466.

[2] "Semi-nilpotent ideals," *Duke Mathematical Journal*, vol. 10 (1943), pp. 553-556.

N. H. McCoy:

[1] "Subrings of direct sums," *American Journal of Mathematics*, vol. LX (1938), pp. 374-382.

[2] "Subrings of infinite direct sums," *Duke Mathematical Journal*, vol. 4 (1938), pp. 486-494.

N. H. McCoy and D. Montgomery:

[1] "A representation of generalized Boolean rings," *Duke Mathematical Journal*, vol. 3 (1937), pp. 455-459.

J. von Neumann:

[1] "On regular rings," *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 707-713.

S. Perlis:

[1] "A characterization of the radical of an algebra," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 128-132.

M. H. Stone:

[1] "The theory of representations for Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.

ON EVEN-DIMENSIONAL SKEW-METRIC SPACES AND THEIR GROUPS OF TRANSFORMATIONS.*

By HWA-CHUNG LEE.

In a previous paper[1] we have studied the differential geometry of an even-dimensional space L_{2n} with a non-singular skew-symmetric fundamental tensor $a_{\alpha\beta}$ $(\alpha, \beta, \gamma, \rho, \sigma, \tau = 1, \cdots, 2n)$. We shall here discuss certain properties of the following transformation groups in such a space.

In the group G of *analytic point-transformations* of L_{2n} there is a subgroup G_1 of *conformal transformations* each leaving the fundamental tensor invariant save for a non-vanishing factor. Referring L_{2n} to an arbitrary chosen coordinate system a conformal point transformation $x \rightarrow y$ satisfies, by definition, the equation

$$\psi a_{\alpha\beta}(x) = a_{\rho\sigma}(y) \frac{\partial y^\rho}{\partial x^\alpha} \frac{\partial y^\sigma}{\partial x^\beta} \qquad (\psi \neq 0). \tag{1}$$

If the factor ψ is restricted to be constant, we have the smaller subgroup G_2 of *special conformal transformations*. If we require $\psi = 1$, we obtain the further subgroup G_3 of *automorphisms*. Thus

$$G_3 \subset G_2 \subset G_1 \subset G.$$

1. Let us first dispose of the case of an L_2. In this case (1) reduces to the single equation

$$\psi a_{12}(x) = a_{12}(y) \left(\frac{\partial y^1}{\partial x^1} \frac{\partial y^2}{\partial x^2} - \frac{\partial y^2}{\partial x^1} \frac{\partial y^1}{\partial x^2} \right).$$

Since every L_2 is flat[2] we may refer it to a coordinate system in which each component of the fundamental tensor is constant,[3] thus $a_{12}(x) = a_{12}(y) =$ const. $(\neq 0)$, so that

$$\psi = \frac{\partial(y^1, y^2)}{\partial(x^1, x^2)}. \tag{2}$$

Hence

THEOREM 1. *For an L_2, the group G_3 of automorphisms is isomorphic*

---

* Received July 24, 1944.

[1] H. C. Lee, "A kind of even-dimensional differential geometry and its application to exterior calculus," *American Journal of Mathematics*, vol. 65 (1943), p. 433.

[2] *Loc. cit.*, p. 434.

[3] *Loc. cit.*, p. 434.

to the unimodular group, and the special conformal group G_2 is isomorphic to the group of transformations of constant functional determinants. The conformal group G_1 is identical with the group G of all transformations, the factor ψ being entirely unrestricted.

We shall supose $2n > 2$. If we differentiate (1) with respect to x^γ, permute α, β, γ cyclically and add, we find

$$(3)\quad \psi K_{\alpha\beta\gamma}(x) + \frac{\partial\psi}{\partial x^\alpha} a_{\beta\gamma}(x) + \frac{\partial\psi}{\partial x^\beta} a_{\gamma\alpha}(x) + \frac{\partial\psi}{\partial x^\gamma} a_{\alpha\beta}(x) = K_{\rho\sigma\tau}(y) \frac{\partial y^\rho}{\partial x^\alpha} \frac{\partial y^\sigma}{\partial x^\beta} \frac{\partial y^\tau}{\partial x^\gamma}$$

where $K_{\alpha\beta\gamma}$ denotes the curvature tensor.[4] Transvecting (3) by the inverse $a^{\beta\gamma}$ of the fundamental tensor, the result is reducible, in consequence of (1), to the form

$$(4)\qquad K_\alpha(x) - (2n-2) \frac{\partial \log \psi}{\partial x^\alpha} = K_\rho(y) \frac{\partial y^\rho}{\partial x^\alpha}$$

where K_α denotes the curvature vector.[5] If we differentiate (4) with respect to x^β, interchange α, β and subtract, we get

$$(5)\qquad b_{\alpha\beta}(x) = b_{\rho\sigma}(y) \frac{\partial y^\rho}{\partial x^\alpha} \frac{\partial y^\sigma}{\partial x^\beta}$$

where $b_{\alpha\beta}$ denotes the first conformal curvature tensor.[6] (5) is a necessary and sufficient condition for (4) to be integrable when regarded as a differential equation in ψ.

If L_{2n} is conformally flat,[7] the first conformal curvature tensor vanishes[8] so that (5) is satisfied identically. Thus (4) is integrable in ψ. In fact in this case a multiplier ϕ exists (determined save for a constant factor) such that $\phi a_{\alpha\beta}$ is the fundamental tensor of the corresponding flat space, and K_α (being a gradient) is expressed in terms of this ϕ by[9]

$$K_\alpha(x) = (2n-2) \frac{\partial \log \phi(x)}{\partial x^\alpha},$$

whence (4) yields

$$(6)\qquad \psi = \text{const.}\ \frac{\phi(x)}{\phi(y)}.$$

Hence

THEOREM 2. *In a conformally flat space L_{2n} $(2n > 2)$ with fundamental*

---

[4] *Loc. cit.*, p. 433.

[5] *Loc. cit.*, p. 434.

[6] *Loc. cit.*, p. 435.

[7] *Loc. cit.*, p. 436.

[8] *Loc. cit.*, p. 436.

[9] *Loc. cit.*, p. 436.

tensor $a_{\alpha\beta}$, *the factor* ψ *of a conformal point-transformation* $x \to y$ *is determined save for a multiplicative constant:*

$$\psi = C \frac{\phi(x)}{\phi(y)},$$

where ϕ *is the multiplier (determined to within a constant factor) which makes* $\phi a_{\alpha\beta}$ *the fundamental tensor of the corresponding flat space.*

In this conformally flat case, $a_{\alpha\beta}$ is expressible in terms of ϕ and of a quantity a_α (determined save for an additive gradient) in the form[10]

$$\phi(x) a_{\alpha\beta}(x) = \frac{\partial}{\partial x^\alpha} a_\beta(x) - \frac{\partial}{\partial x^\beta} a_\alpha(x).$$

We can then reduce (1) to

$$\frac{\partial}{\partial x^\alpha} \left\{ C a_\beta(x) - a_\sigma(y) \frac{\partial y^\sigma}{\partial x^\beta} \right\} - \frac{\partial}{\partial x^\beta} \left\{ C a_\alpha(x) - a_\rho(y) \frac{\partial y^\rho}{\partial x^\alpha} \right\} = 0,$$

whence

$$C a_\alpha(x) - a_\rho(y) \frac{\partial y^\rho}{\partial x^\alpha} = \text{gradient}.$$

The gradient on the right may be dropped on account of the additive gradient in a_α. Hence

THEOREM 3. *The conformal group* G_1 *of a conformally flat space* L_{2n} $(2n > 2)$ *consists of those point-transformations under which* a_α *behaves like a covariant pseudovector (vector except for a multiplicative constant):*

$$C a_\alpha(x) = a_\rho(y) \frac{\partial y^\rho}{\partial x^\alpha},$$

where a_α *is the quantity (determined to within an additive gradient) in terms of which the fundamental tensor of the corresponding flat space is expressed in the form*

$$\frac{\partial}{\partial x^\alpha} a_\beta(x) - \frac{\partial}{\partial x^\beta} a_\alpha(x).$$

If L_{2n} itself is flat, then $\phi(x) = \phi(y) = \text{const.}$ so that (6) implies $\psi = \text{const.}$ Conversely $\psi = \text{const.}$ implies also $\phi = \text{const.}$ and so L_{2n} is flat. Thus

THEOREM 4. *For a flat space* L_{2n} $(2n > 2)$ *the special conformal group*

---

[10] *Loc. cit.*, p. 436.

G_2 is identical with the conformal group G_1. Conversely, if for a conformally flat space L_{2n} $(2n > 2)$ G_2 is identical with G_1, the space itself is flat.

Before identifying the group G_3 of automorphisms of a flat L_{2n} we first introduce the generalized Poisson parentheses.

2. Consider in a general L_{2n} two functions of position and form the expression

$$(F_1, F_2) = a^{\alpha\beta} \frac{\partial F_1}{\partial x^\alpha} \frac{\partial F_2}{\partial x^\beta} . \tag{7}$$

Evidently we have

$$(F_1, F_2) = -(F_2, F_1),$$

and it is also easy to verify the identity

$$(F_1, (F_2, F_3)) + (F_2, (F_3, F_1)) + (F_3, (F_1, F_2)) = K^{\alpha\beta\gamma} \frac{\partial F_1}{\partial x^\alpha} \frac{\partial F_2}{\partial x^\beta} \frac{\partial F_3}{\partial x^\gamma}$$

where $K^{\alpha\beta\gamma}$ denotes the contravariant associate of the curvature tensor:

$$K^{\alpha\beta\gamma} = a^{\alpha\rho} a^{\beta\sigma} a^{\gamma\tau} K_{\rho\sigma\tau} = a^{\alpha\rho} \frac{\partial a^{\beta\gamma}}{\partial x^\rho} + a^{\beta\sigma} \frac{\partial a^{\gamma\alpha}}{\partial x^\sigma} + a^{\gamma\tau} \frac{\partial a^{\alpha\beta}}{\partial x^\tau} .$$

The above identity reduces to the Jacobi relation

$$(F_1, (F_2, F_3)) + (F_2, (F_3, F_1)) + (F_3, (F_1, F_2)) = 0$$

when L_{2n} is flat, referred to any coordinate system whatever.

For a flat L_{2n}, there exist coordinate systems in which the components of the fundamental tensor are constants.[11] Such coordinate systems may be called *preferred.* In particular a preferred coordinate system will be called *canonical* when

$$a_{\alpha\beta} = \begin{pmatrix} O & I \\ -I & O \end{pmatrix}, \qquad a^{\alpha\beta} = \begin{pmatrix} O & -I \\ I & O \end{pmatrix}, \tag{8}$$

where I and O denote the unit and zero matrices of order n respectively.

In a flat L_{2n} referred to a canonical coordinate system, (7), on account of (8), may be written in the form

$$(F_1, F_2) = \sum_{i=1}^{n} \left\{ \frac{\partial F_1}{\partial p_i} \frac{\partial F_2}{\partial x^i} - \frac{\partial F_1}{\partial x^i} \frac{\partial F_2}{\partial p_i} \right\} \tag{9}$$

where $p_i = x^{n+i}$ $(i = 1, \cdots, n)$. This is the well known *Poisson parenthesis.*

To consider the group G_3 of automorphisms of a flat L_{2n} put $\psi = 1$ in (1) and write it in the equivalent form

---

[11] *Loc. cit.*, p. 434.

$$a^{\rho\sigma}(y) = a^{\alpha\beta}(x)\,\frac{\partial y^\rho}{\partial x^\alpha}\,\frac{\partial y^\sigma}{\partial x^\beta}. \tag{10}$$

The right-hand side, in canonical coordinates, is no other than the Poisson parenthesis (y^ρ, y^σ). Thus (10) in these coordinates decomposes into

$$(y^i, y^j) = 0, \quad (q_i, y^j) = \delta_i{}^j, \quad (q_i, q_j) = 0, \qquad (i, j = 1, \cdots, n) \tag{11}$$

where $q_i = y^{n+i}$, which are precisely *Poisson's conditions* for a *canonical transformation*[12] or in particular *for a homogeneous contact transformation.*[13] Hence

THEOREM 5. *The group G_3 of automorphisms of a flat L_{2n} is isomorphic to the group of canonical transformations in 2n variables.*

We remark that (10) is equivalent to

$$a_{\alpha\beta}(x) = a_{\rho\sigma}(y)\,\frac{\partial y^\rho}{\partial x^\alpha}\,\frac{\partial y^\sigma}{\partial x^\beta}$$

(which is just (1) with $\psi = 1$) whose right-hand side is, by (8), of the form

$$\sum_{i=1}^{n}\left\{\frac{\partial y^i}{\partial x^\alpha}\,\frac{\partial q_i}{\partial x^\beta} - \frac{\partial q_i}{\partial x^\alpha}\,\frac{\partial y^i}{\partial x^\beta}\right\}$$

which is *Lagrange's parenthesis*. Thus the above equation in a canonical coordinate system represents *Lagrange's condition* for a canonical transformation. Similarly, (10) is also equivalent to each of the following two

$$a^{\alpha\beta}(x) = a^{\rho\sigma}(y)\,\frac{\partial x^\alpha}{\partial y^\rho}\,\frac{\partial x^\beta}{\partial y^\sigma}, \qquad a_{\rho\sigma}(y) = a_{\alpha\beta}(x)\,\frac{\partial x^\alpha}{\partial y^\rho}\,\frac{\partial x^\beta}{\partial y^\sigma},$$

which in a canonical coordinate system represent the Poisson and Lagrange conditions respectively for the inverse transformation to be a canonical transformation.

3. The special conformal group G_2 of any L_{2n} has a very remarkable property in the following connection. Consider a system of curves in L_{2n} defined by a system of ordinary differential equations of the form

$$\frac{dx^\alpha}{dt} + a^{\alpha\beta}(x)\,\frac{\partial H(x, t)}{\partial x^\beta} = 0\,. \tag{12}$$

---

[12] See, for example, E. Goursat, *Leçons sur l'intégration des équations aux dérivées partielles du premier ordre*, 2nd ed. (1921), p. 387.

[13] See, for example, L. P. Eisenhart, *Continuous groups of transformations* (1933), p. 250.

where t is a parameter and H is a given function of t and of the x's. We call such a system of curves a *Hamiltonian congruence*, and the function H the *Hamiltonian* of the congruence. These denominations are justified when L_{2n} is flat and is referred to a canonical coordinate system, for in this case (12) may be written on account of (8) in the form

$$\frac{dx^i}{dt}=\frac{\partial H}{\partial p_i},\quad \frac{dp_i}{dt}=-\frac{\partial H}{\partial x^i},\qquad (i=1,\cdots,n)$$

which are the well known Hamilton's canonical equations in Analytical Mechanics. By a point-transformation in L_{2n}, a Hamiltonian congruence of curves is transformed into a system of curves which is in general no longer a Hamiltonian congruence. We call a point transformation a *Hamiltonian transformation* if it transforms *every* Hamiltonian congruence into a Hamiltonian congruence. The Hamiltonian transformations in L_{2n} evidently form a group, which we shall now show to be no other than the special conformal group G_2. In other words

THEOREM 6. *The special conformal group G_2 of any space L_{2n} consists of the Hamiltonian transformations of the space.*

To prove this theorem consider a Hamiltonian transformation $x\rightarrow y$ which, by definition, transforms (12) for an arbitrary H into an equation of the same form

$$\frac{dy^\rho}{dt}+a^{\rho\sigma}(y)\frac{\partial\tilde{H}(y,t)}{\partial y^\sigma}=0. \tag{13}$$

Since

$$\frac{dy^\rho}{dt}=\frac{\partial y^\rho}{\partial x^\alpha}\frac{dx^\alpha}{dt}=-\frac{\partial y^\rho}{\partial x^\alpha}a^{\alpha\beta}(x)\frac{\partial H(x,t)}{\partial x^\beta}$$

by (12), (13) may be written

$$\frac{\partial\tilde{H}}{\partial y^\sigma}=a_{\sigma\rho}(y)\frac{\partial y^\rho}{\partial x^\alpha}a^{\alpha\beta}(x)\frac{\partial H}{\partial x^\beta}. \tag{14}$$

To obtain the conditions of integrability of (14) for $\tilde{H}$ we first find by differentiation

$$\frac{\partial}{\partial y^\tau}\left\{\frac{\partial\tilde{H}}{\partial y^\sigma}\right\}=\frac{\partial}{\partial y^\tau}\left\{a_{\sigma\rho}(y)\frac{\partial y^\rho}{\partial x^\alpha}a^{\alpha\beta}(x)\right\}\frac{\partial H}{\partial x^\beta}+a_{\sigma\rho}(y)\frac{\partial y^\rho}{\partial x^\alpha}a^{\alpha\beta}(x)\frac{\partial x^\gamma}{\partial y^\tau}\frac{\partial}{\partial x^\gamma}\left\{\frac{\partial H}{\partial x^\beta}\right\}.$$

Subtracting from this equation the equation obtained from it by interchanging

σ and τ, and in the result putting the coefficients of $\partial H/\partial x^\beta$ and $\partial^2 H/\partial x^\gamma \partial x^\beta$ (which latter is symmetric in β, γ) equal to zero (since H is arbitrary), we obtain the required conditions

$$(15) \qquad \frac{\partial}{\partial y^\tau}\left\{ a_{\sigma\rho}(y) \frac{\partial y^\rho}{\partial x^\alpha} a^{\alpha\beta}(x) \right\} - \frac{\partial}{\partial y^\sigma}\left\{ a_{\tau\rho}(y) \frac{\partial y^\rho}{\partial x^\alpha} a^{\alpha\beta}(x) \right\} = 0,$$

$$(16) \quad a_{\sigma\rho}(y)\frac{\partial y^\rho}{\partial x^\alpha}\left\{ a^{\alpha\beta}(x)\frac{\partial x^\gamma}{\partial y^\tau} + a^{\alpha\gamma}(x)\frac{\partial x^\beta}{\partial y^\tau} \right\} - a_{\tau\rho}(y)\frac{\partial y^\rho}{\partial x^\alpha}\left\{ a^{\alpha\beta}(x)\frac{\partial x^\gamma}{\partial y^\sigma} + a^{\alpha\gamma}(x)\frac{\partial x^\beta}{\partial y^\sigma} \right\} = 0.$$

To simplify these conditions we first transvect (16) by $\dfrac{\partial y^\sigma}{\partial x^\lambda}\dfrac{\partial y^\tau}{\partial x^\mu}$ $(\lambda, \mu = 1, \cdots, 2n)$, finding that (16) is equivalent to

$$(17) \quad w_{\lambda\alpha}\{a^{\alpha\beta}(x)\delta_\mu{}^\gamma + a^{\alpha\gamma}(x)\delta_\mu{}^\beta\} - w_{\mu\alpha}\{a^{\alpha\beta}(x)\delta_\lambda{}^\gamma + a^{\alpha\gamma}(x)\delta_\lambda{}^\beta\} = 0$$

where for the moment we put

$$(18) \qquad w_{\alpha\beta} = a_{\rho\sigma}(y)\frac{\partial y^\rho}{\partial x^\alpha}\frac{\partial y^\sigma}{\partial x^\beta}.$$

Contracting (17) for γ and μ we find

$$w_{\lambda\alpha}a^{\alpha\beta}(x) = \psi\delta_\lambda{}^\beta \qquad (\psi = - w_{\alpha\beta}a^{\alpha\beta}(x)/2n)$$

which may be written

$$(19) \qquad w_{\alpha\beta} = \psi a_{\alpha\beta}(x).$$

Since this satisfies (17), it is equivalent to (16). Replacing $w_{\alpha\beta}$ in (19) by its definition (18), (19) assumes the form

$$(20) \qquad \psi a_{\alpha\beta}(x) = a_{\rho\sigma}(y)\frac{\partial y^\rho}{\partial x^\alpha}\frac{\partial y^\sigma}{\partial x^\beta}.$$

Because of (20) we may write (15) thus

$$\frac{\partial}{\partial y^\tau}\left\{ \psi\frac{\partial x^\beta}{\partial y^\sigma} \right\} - \frac{\partial}{\partial y^\sigma}\left\{ \psi\frac{\partial x^\beta}{\partial y^\tau} \right\} = 0$$

which implies

$$(21) \qquad \psi = \text{const.}$$

We see from (20) and (21) that the Hamiltonian transformation $x \to y$ in question is a special conformal transformation.

Consider, conversely, a special conformal transformation $x \to y$. It satisfies (20) and (21) by definition. Since the latter are equivalent to (15) and (16), they are, therefore, the conditions of integrability of (14) for $\tilde{H}$.

Hence (14) furnishes an $\bar{H}$ corresponding to an arbitrarily given H. By means of (14), (12) is transformed into (13), whence the special conformal transformation $x \rightarrow y$ is a Hamiltonian transformation. Theorem 6 is proved.

In consequence of (20) we may write (14) in the form

$$\frac{\partial \bar{H}}{\partial x^{\alpha}} = \psi \frac{\partial H}{\partial x^{\alpha}}$$

whence

$$\bar{H}(y, t) = \psi H(x, t) + \chi(t). \tag{22}$$

The additive term $\chi(t)$, which is an arbitrary function of t, may be omitted, since a Hamiltonian may be increased by any function of t without affecting the corresponding Hamiltonian congruence. Hence

THEOREM 7. *When a Hamiltonian congruence is transformed by a special conformal transformation of factor ψ, the Hamiltonian of the congruence is simply multiplied by this factor ψ.*

ACADEMIA SINICA.

CALCUTTA UNIVERSITY
CENTRAL LIBRARY.

A DUAL-SYMMETRIC DEFINITION OF FIELD.*

By Alfred L. Foster and B. A. Bernstein.

1. Introduction. In another paper [1][1] it was shown how the general theory of (commutative) rings (with unit) may be developed in terms of operations symmetrically related to each other, as opposed to the traditional asymmetrical $+$, $\times$ treatment. It was shown that this symmetrical approach discloses a *duality theorem for rings* which, in particular, subsumes the familiar duality theorem of Boolean Algebra. In this dual-symmetric approach to rings initiated in [1] the following *dual pairs*; $+, \underset{\circ}{+}$; $\times, \Delta$; as well as the (self-dual) *ring complement*, $*$, were shown to play a central role, where

$$a + b = \textit{ordinary ring addition.}$$
$$a \underset{\circ}{+} b = a + b - 1.$$
$$a \times b = \textit{ordinary ring multiplication (also written } ab\textit{).}$$
$$a \,\Delta\, b = a + b - ab, \textit{ called ``ring } \Delta\textit{-multiplication.''}$$

It is sufficient here to recall a considerably curtailed version of the ring-duality theorem given in [1]. Let $a, b \cdots$ denote variable ring elements.

Ring-Duality Theorem. *If* $P(a, b, \cdots; +, \underset{\circ}{+}; \times, \Delta; *; 0, 1)$ *is a true ring proposition, so also is the dual proposition,* $dl\, P$, *obtained by replacing*

$+$ *by* $\underset{\circ}{+}$
$\underset{\circ}{+}$ *by* $+$
$\times$ *by* Δ
Δ *by* $\times$
0 *by* 1
1 *by* 0
$*$ *by* $*$ (*i. e., unchanged*).

---

* Received May 8, 1944. Presented to the American Mathematical Society, March, 1944.

[1] "Symmetric approach to commutative rings, with duality theorem; Boolean duality as special case," by A. L. Foster and B. A. Bernstein, *Duke Mathematical Journal*, vol. 11 (1944), pp. 603-616.

The present paper, the second to deal with the development of the theory of rings from the dual-symmetric approach, may be read with only the above mentioned knowledge of [1]. We here give a symmetric definition of the concept of (commutative) field entirely in terms of its *dual multiplications* $\times$ and Δ. For a field (or for any species of ring) so characterized, the ring-duality theorem obviously assumes the extremely simple form:

$$dl\,P(a, b, \cdots; \times, \Delta, 0, 1) = P(a, b, \cdots; \Delta, \times, 1, 0).$$

Two sets (I and II) of postulates for the concept field are given, and moreover in each set the postulates occur in dual pairs; and therefore an independent direct proof of the duality theorem (only, of course, for fields) follows by inspection from (either set of) these postulates. In set I the definition of $+$ (in terms of the primitive notions F, $\times$, Δ, 0, 1) is given in **4**, and it is shown that with this $+$ the system $(F, +, \times)$ is (strongly) equivalent to the usual definition of field. For the most part only such theorems are proved as bear directly on this question of equivalence, though a few (of many) theorems belonging to this dual-symmetric theory have been included for their own intrinsic beauty of form. The set II is taken up in **6**.

The independence of the postulates has not been fully established, though each set, with the possible exception of the "-1 postulate" (i. e., D of set I, $=$ P7 of set II) is believed independent when appropriate restrictive clauses are introduced. In this connection it will be seen that the -1 postulate is self-dual, and hence the postulates (of either set) which remain after its deletion, (in the event of its non-independency), still occur in dual pairs.

2. The postulates, Set I. As primitive notions in the dual-symmetric definition of field to be given we take, then, an undefined class F and two undefined binary operations, $\times$ and Δ; and for convenience of formulation we also take 0 and 1 as undefined elements of F. Our primitive system is then

$$(F, \times, \Delta, 1, 0).$$

Furthermore numerous cumbersome special cases in the proofs are avoided if we assume, as we henceforth shall, that we are defining fields of at least three elements.

Notation. F_ϕ denotes the class F after deleting the element 0; similarly F_1 denotes F after deleting the element 1. Frequently $a \times b$ is abbreviated to ab. The inverse of an element a with respect to the group $\times$ (postulated below) is denoted by a', and the inverse of a with respect to the group Δ is denoted by a°. Also a'° denotes $(a')^\circ$, and similarly for $a^{\circ\prime}$, $a^{\circ\prime\circ}$, etc. Small

Roman letters a, b, etc. denote elements of F. Finally for a, b, c belonging to the group $\times$ (respectively to the group Δ), we write $(ab)c = a(bc) = abc$, (respectively $a \Delta (b \Delta c) = (a \Delta b) \Delta c = a \Delta b \Delta c$).

The postulates of set I are:

A. *0 and 1 are elements of F and F contains an element* $\neq 0, \neq 1$.

(*Field*) *Group Postulates*

B (i). $(F_{\phi}, \times)$ *is an Abelian group with unit element* 1.
(ii). $0a = a0 = 0$.

B′ (i). (F_{+}, Δ) *is an Abelian group with unit element* 0.
(ii). $1 \Delta a = a \Delta 1 = 1$.

(*Field*) *Linking Postulates*

C. $ab = a \Delta (a' \Delta b^{\circ})'^{\circ}$ $[a \neq 0, \neq 1;\ b \neq 1;\ a' \Delta b^{\circ} \neq 0$

C′. $a \Delta b = a(a^{\circ}b')^{\circ\prime}$ $[a \neq 0, \neq 1;\ b \neq 0;\ a^{\circ}b' \neq 1]$

D. $(aa^{\circ\prime}a'^{\circ})^{\circ} = (b \Delta b'^{\circ} \Delta b^{\circ\prime})'$ $[a \neq 0, \neq 1;\ b \neq 0, \neq 1]$.

It is easily verified that with $\Delta, \times, 0, 1$ interpreted as stated (for rings) in 1, each of the postulates is a true proposition of the concept field $(F, +, \times, 0, 1)$ as ordinarily defined.

Concerning the use of the = sign. In the postulates, and in the theorems to follow, the = sign, except where otherwise explicitly stated, is used in a formula

$$S(a, b, \cdots) = T(a, b, \cdots) \tag{1}$$

always in the sense: the two sides take on the same value for all values of the arguments $a, b, \cdots$ belonging to the *domain* of the formula, i. e., for all values of $a, b, \cdots$ for which both sides are defined. Frequently this convention is augmented by an explicit statement, in square [ ] brackets, of the domain of the formula. The domain of one formula, (1), (or simply of a function $S(a, b, \cdots)$) is *included* ($\subset$) in that of another formula (respectively function) if for all values of the arguments for which the first is defined the second is also. With this use of = certain elementary precautions are necessary in the derivation of theorems. We mention several frequently used obvious facts about class inclusion as applied to domains; in each case it is merely a question of establishing that the domain of a formula is not increased by the indicated transformations.

Σ1. If $S(a, b, \cdots) = T(a, b, \cdots)$ is a theorem (or postulate), and if U is a function, then each of: $S' = T'$; $S^{\circ} = T^{\circ}$; $U \times S = U \times T$; $U \Delta S = U \Delta T$ is also a theorem.

Σ2. If $S(\cdots, b, \cdots) = T(\cdots, b, \cdots)$ is a theorem so are $S(\cdots, b', \cdots) = T(\cdots, b', \cdots)$ and $S(\cdots, b^\circ, \cdots) = T(\cdots, b^\circ, \cdots)$.

Σ3. If the expression U'' (respectively $U^{\circ\circ}$) occurs in a theorem, and if the domain of the formula which results from the replacement of U'' by U (respectively of $U^{\circ\circ}$ by U) is included in the domain of the theorem, then the resulting formula is a theorem. The same holds with respect to the replacement of the expression $U \times U'$ by 1 (respectively of $U \Delta U^\circ$ by 0).

These Σ theorems will be employed without further mention.

Concerning Duality. As already observed the postulates are seen to occur in dual pairs: A and D each being self-dual; B, B′; C, C′, where the dual of any postulate is obtained by replacing

$$\begin{array}{ccc} \times & \text{by} & \Delta \\ \Delta & \text{``} & \times \\ 0 & \text{``} & 1 \\ 1 & \text{``} & 0. \end{array} \tag{2}$$

This of course implies that the inverses, ', ° of $\times$ and Δ must be replaced

$$\begin{array}{ccc} {}' & \text{by} & {}^\circ \\ {}^\circ & \text{``} & {}' \end{array} \tag{2.1}$$

in obtaining the dual. In view of this the duality theorem (for fields) follows immediately from these postulates. Furthermore in proving any theorem we can dualize the proof and thereby obtain a proof for the dual theorem. It is therefore only necessary to prove one of a dual pair T, T' of theorems,—the other may then simply be written down by means of the exchanges (2) and (2.1).

3. General theorems. For convenient reference we give three immediate consequences of A, B, B′.

T1. (= THEOREM 1). $0 \neq 1$, *and F contains at least three elements.*

T2.
$$\begin{aligned} 0 \Delta a &= a \Delta 0 = a \\ 1a &= a1 = a \\ 0a &= a0 = 0 \\ 1 \Delta a &= a \Delta 1 = 1. \end{aligned}$$

T3.
$$\begin{aligned} 1' &= 1 \\ 0^\circ &= 0 \\ a'' &= a \\ a^{\circ\circ} &= a \\ a' &= b' \rightleftarrows a = b \end{aligned}$$

$$\begin{aligned}
a^{\circ} &= b^{\circ} \rightleftarrows a = b \\
a &\neq 0 \rightarrow a' \text{ exists and } \neq 0 \\
a &\neq 1 \rightarrow a^{\circ} \quad \text{“} \quad \text{“} \quad \neq 1 \\
a &\neq 0, \neq 1 \rightarrow a', a^{\circ}, a^{\circ\prime}, a^{\prime\circ}, \neq 0, \neq 1 \\
(ab)' &= a'b' \\
(a \Delta b)^{\circ} &= a^{\circ} \Delta b^{\circ} \\
aa' &= 1 \\
a \Delta a^{\circ} &= 0 \\
ab &= 0 \leftrightarrows a = 0 \text{ or } b = 0 \\
a \Delta b &= 1 \leftrightarrows a = 1 \text{ or } b = 1.
\end{aligned}$$

T4. $aa^{\circ\prime}a^{\prime\circ} = \mu = \textit{constant} \neq 0 \quad [a \neq 0, \neq 1]$.

Proof. For $b \neq 0, \neq 1$

$$b \Delta b^{\prime\circ} \Delta b^{\circ\prime} \neq 1$$

by T3. Suppose that (*Case* 1): an element $b \neq 0, \neq 1$ exists such that

$$b \Delta b^{\prime\circ} \Delta b^{\circ\prime} = \tau \neq 0.$$

Then $\tau \neq 0, \neq 1$, whence $\tau' \neq 0, \neq 1$ (T3). Then from D, for $a \neq 0, \neq 1$

$$(aa^{\circ\prime}a^{\prime\circ})^{\circ} = \tau'$$

and the theorem is proved with $\mu = \tau^{\prime\circ} \neq 0$. (Moreover in this case we have further that $\mu \neq 1$). (*Case* 2): Suppose that for every $b \neq 0, \neq 1$, $b \Delta b^{\prime\circ} \Delta b^{\circ\prime} = 0$. Then from D, for $a \neq 0, \neq 1$ we have $aa^{\circ\prime}a^{\prime\circ} = 1$, and the theorem is proved with $\mu = 1$.

T4′. (= *dual of* T4). $a \Delta a^{\prime\circ} \Delta a^{\circ\prime} = \tau = \textit{constant} \neq 1$, $[a \neq 0, \neq 1]$.

It is seen from T4 and T4′ that the constants μ and τ are *dual* elements, and hence, when obtaining the dual of a field proposition involving them, each must be replaced by the other. It is further easily verified that these constants μ and τ turn out, in a field, to have the realizations “-1” and “$2 (= 1 + 1)$” respectively. (See footnote to T19′). In the proof of T4 we have further shown

T5. (i) $\mu = 1 \rightleftarrows \tau = 0$

(ii) *If* $\mu \neq 1$ (*i. e.*, $\tau \neq 0$), *then* $\mu^{\circ} = \tau'$.

T6. $\mu' = \mu$.

Proof. By T4 $\mu \neq 0$, therefore μ' exists and is given by

$$\begin{aligned}
\mu' &= (aa^{\circ\prime}a^{\prime\circ})' \quad [a \neq 0, \neq 1] \\
&= a'(a^{\circ\prime})'(a^{\prime\circ})', = \mu
\end{aligned}$$

by T's 3, 4.

T6′. $\tau^{\circ} = \tau$.

T7. $aa = 1 \rightarrow a = 1$ *or* $a = \mu$.

Conversely, $a = 1$ *or* $a = \mu : \rightarrow aa = 1$.

Proof. From T4, since $\times$ is a group, $a^{\circ\prime}a^{\prime\circ} \neq 1$ unless $a = \mu$. Hence $a^{\prime\circ} \neq a^{\circ}$ unless $a = \mu$, i. e., $aa \neq 1$ unless $a = \mu$ or $a = 1$ (T's 4, 3, 2). The converse: For $a = 1$ this follows by T2. For $a = \mu \neq 1$,

$$\mu = bb^{\circ\prime}b^{\prime\circ}$$
$$\mu = b^{\prime}b^{\prime\circ\prime}b^{\prime\prime\circ}$$

for $b \neq 0, \neq 1$ (T's 4, 3). Therefore $\mu\mu = bb^{\circ\prime}b^{\prime\circ}b^{\prime}b^{\prime\circ\prime}b^{\circ} = 1$ (T3).

T7′. $a \,\Delta\, a = 0 \rightarrow a = 0$ *or* $a = \tau$.

Conversely, $a = 0$ *or* $a = \tau : \rightarrow a \,\Delta\, a = 0$.

In proving T7 we have also established

T8. *If* $\mu = 1$, *then for all* $a \neq 0, \neq 1$, $a^{\circ\prime}a^{\prime\circ} \neq 1$. *If* $\mu \neq 1$, *then* $a^{\circ\prime}a^{\prime\circ} = 1$ *only for* $a = \mu$.

T8′. *If* $\tau = 0$, *then for all* $a \neq 0, \neq 1$, $a^{\prime\circ} \,\Delta\, a^{\circ\prime} \neq 0$. *If* $\tau \neq 0$, *then* $a^{\prime\circ} \,\Delta\, a^{\circ\prime} = 0$ *only for* $a = \tau$.

T9. $a^{\prime\circ\prime} = a^{\circ\prime\circ}$ $[a \neq 0, \neq 1]$.

Proof. $a^{\prime}a^{\prime\circ\prime}a^{\prime\prime\circ} = \mu$ ($a^{\prime}$ for a in T4), $= a^{\prime}a^{\prime\circ\prime}a^{\circ}$ ($\Sigma 3$).
$a^{\circ}a^{\circ\circ\prime}a^{\circ\prime\circ} = \mu$ (a° “ a “), $= a^{\circ}a^{\prime}a^{\circ\prime\circ}$.

The theorem follows from these two expressions for μ, by T3. It is to be noted that T9 is a self-dual theorem.

We now define a^{*}, the (field) *complement* of a by

Definition E1.

$$a^{*} = \begin{cases} a^{\prime\circ\prime} & \text{if } a \neq 0, \neq 1 \\ 1 & \text{if } a = 0 \\ 0 & \text{if } a = 1. \end{cases}$$

T10.

$$a^{*} = \begin{cases} a^{\circ\prime\circ}\ (= a^{\prime\circ\prime}) & \text{if } a \neq 0, \neq 1 \\ 1 & \text{if } a = 0 \\ 0 & \text{if } a = 1. \end{cases}$$

Proof. T9, E1.

Comparing E1 and T10 it is seen that the operation * is self-dual and hence * is left unchanged in obtaining the dual of any field proposition involving this operation.

T11. *a^* is a unique element of F for each a of F.*

Proof. B, B′, T10.

T12. $a^{**} = a$.

Proof. $0^{**} = 1^* = 0$; $1^{**} = 0^* = 1$; for $a \neq 0, \neq 1$, $a^{**} = a^{\prime\circ\prime\prime\circ\prime} = a$ (T's 10, 3).

T13. $a^* = b^* \leftrightarrow a = b$

$a^* = b \leftrightarrow a = b^*$.

Proof. T's 11, 12.

T14. $a^{\prime\circ} = a^{*\prime}$ $[a \neq 0, \neq 1]$.

Proof. $a^{*\prime} = a^{\prime\circ\prime\prime} = a^{\prime\circ}$ (T's 10, 3).

T14′. $a^{\circ\prime} = a^{*\circ}$ $[a \neq 0, \neq 1]$.

T15. $a^{\prime\circ} = a^{\circ *}$ $[a \neq 0, \neq 1]$.

Proof. $a^{\circ *} = a^{\circ\circ\prime\circ} = a^{\prime\circ}$ (T's 10, 3).

T15′. $a^{\circ\prime} = a^{\prime *}$ $[a \neq 0, \neq 1]$.

T16. $a^{\circ *} = a^{*\prime}$ $[a \neq 1]$.

Proof. $0^{\circ *} = 0^* = 1$; $0^{*\prime} = 1' = 1$ (T's 3, 10); for $a \neq 0, \neq 1$, $a^{\circ *} = a^{\prime\circ}, = a^{*\prime}$ (T's 10, 14, 3).

T16′. $a^{\prime *} = a^{*\circ}$ $[a \neq 0]$.

From Theorems 14-16′ and T's 4, 4′ one easily culls the following theorem, symmetric in the three unary operations $'$, $^\circ$, * which we call the

T17 *Unary Operations Theorem.*[2] *Let $\sigma(=\sigma_1), \sigma_2, \sigma_3$ denote (in any desired order) the three unary operations $'$, $^\circ$, * and let a, b be $\neq 0, \neq 1$. Then*

(i) $a^\sigma \neq 0, \neq 1$

(ii) $a = b \rightleftarrows a^\sigma = b^\sigma$

---

[2] In [1] it is shown that in the case of a general ring, R, Theorem 17 holds in $\mathfrak{U}$, the *double-unity* subset of R, i. e., the subset of those elements of R which are unity elements both with respect to the group $\times$ and with respect to its dual, Δ.

$$
\begin{aligned}
&\text{(iii)} \quad a^{\sigma\sigma} = a \\
&\text{(iv)} \quad a^{\sigma_1\sigma_2\sigma_3} = a^{\sigma_2} \\
&\text{(v)} \quad a^{\sigma_1\sigma_2\sigma_1} = a^{\sigma_3} \\
&\text{(vi)} \quad a^{\sigma_1\sigma_2} = a^{\sigma_2\sigma_3} \\
&\text{(vii)} \quad a^{\sigma_1}a^{\sigma_2}a^{\sigma_3} = \mu \\
&\text{(viii)} \quad a^{\sigma_1} \Delta a^{\sigma_2} \Delta a^{\sigma_3} = \tau.
\end{aligned}
$$

By comparison with the definition E1, (v) above shows that for $a \neq 0, \neq 1$ any one of the three unary operations $^\circ$, $'$, * is definable in terms of the other two, and moreover in each instance by equations of the same form, namely of the form given in E1.

T18. $\tau = \mu^*$
$\mu = \tau^*$.

Proof. The second formula follows from the first by T12. Consider then the first, $\tau = \mu^*$. If $\tau = 0$ then $\mu = 1$, and the theorem is immediate, (T's 5, 10). For $\tau \neq 0$ (i. e., $\mu \neq 1$, by T5), $\mu^\circ = \tau'$ (D, T's 4, 4′). Therefore $\tau = \tau^\circ = \mu^{\circ\prime\circ} = \mu^*$ (T's 6′, 3, 10).

T19. $\mu^{*\circ} = \mu^* = \tau$.

Proof. T's 18, 6′.

T19′.[3] $\tau^{*\prime} = \tau^* = \mu$.

T20. $ab = a \Delta a^\circ b^*$ $[a \neq 1]$.

Proof. Putting $a^\circ b' = q^{\prime\circ}$, $b' = a^{\circ\prime} q^{\prime\circ}$ in C′, we get

$$(1) \qquad a \Delta a^{\circ\prime\prime} q^{\prime\circ\prime} = a q^{\prime\circ\circ\prime} \qquad [a \neq 0, \neq 1;\ a^{\circ\prime\prime} q^{\prime\circ\prime} \neq 0;\ q^{\prime\circ} \neq 1].$$

Using T's 3, 10, (1) establishes

---

[3] $\mu = 1$ (i. e., $\tau = 0$) only for fields of characteristic 2. For fields of characteristic $\neq 2$ the *center* (element), ω, plays a more symmetrical role than either μ or τ, where *Definition.*

$$\omega = \begin{cases} 1 \text{ for a field of characteristic 2.} \\ \mu^\circ [= \tau' = (a a^{\circ\prime} a^{\prime\circ})^\circ = (a \Delta a^{\prime\circ} \Delta a^{\circ\prime})'] \\ \text{for a field of characteristic } \neq 2. \quad \text{(Compare with postulate D).} \end{cases}$$

One then easily has the

THEOREM. *For fields of characteristic* $\neq 2$ $\omega^\circ = \mu$, $\omega' = \tau$, $\omega^* = \omega$.

The center of a field of characteristic p is readily computed to be given by

$$\omega \begin{cases} = \tfrac{1}{2} (= (1+1)') \text{ if } p = 0 \\ = (p-1)/2 \text{ (i. e., } \underbrace{1 + 1 + \cdots + 1}_{(p-1)/2 \text{ terms}}) \text{ if } p \neq 0. \end{cases}$$

$$(2)\qquad a \,\Delta\, a^{\circ}q^{*} = aq \quad (\text{for } a \neq 0, \neq 1;\ q \neq 0, \neq 1).$$

The domain of (2), however, unlike that of (1), also permits $a = 0$, $q = 0$, or $q = 1$. For each of these cases (2) is immediately true by T's 3, 2 and E1, which establishes T20.

T20′. $a \,\Delta\, b = a(a' \,\Delta\, b^{*})$ $\quad [a \neq 0]$.

T21. $(a \,\Delta\, b)^{*} = a^{*}b^{*}$.

Proof. By T20 and C we have respectively

$$ab = a \,\Delta\, a^{\circ}b^{*} \qquad [a \neq 1].$$

$$ab = a \,\Delta\, (a' \,\Delta\, b^{\circ})^{\prime\circ} \qquad [a \neq 0, \neq 1;\ b \neq 1;\ a' \,\Delta\, b^{\circ} \neq 0].$$

Since ab is defined for all a, b of F, we then have, by T3,

$$\begin{aligned} a^{\circ}b^{*} &= (a' \,\Delta\, b^{\circ})^{\prime\circ} && [a \neq 0, \neq 1;\ a' \,\Delta\, b^{\circ} \neq 0] \\ &= (a^{\prime\circ} \,\Delta\, b)^{\circ\prime\circ} && [\quad ``\qquad\quad ``\quad ``\quad]. \end{aligned}$$

Putting $a^{\prime\circ} = q$ and using T10 this becomes

$$(1)\qquad q^{*}b^{*} = (q \,\Delta\, b)^{*} \quad (\text{for } q \neq 0, \neq 1;\ b \neq 1;\ q \,\Delta\, b \neq 0).$$

The domain of (1), however, permits all values of q and b. For each of the cases $q = 0$, $q = 1$, or $b = 1$, (1) is directly verified by T's 2, 10. And for the case $q \,\Delta\, b = 0$ (1) is also true, since

$$q \,\Delta\, b = 0 (q \neq 0, \neq 1;\ b \neq 0, \neq 1) \begin{cases} \rightarrow (q \,\Delta\, b)^{*} = 1 \quad (\text{E1}). \\ \rightarrow q = b^{\circ}, \rightarrow q^{\prime\circ} = b^{\circ\prime\circ}, \rightarrow q^{\prime\circ\prime}b^{\circ\prime\circ} \\ \qquad = 1, \rightarrow q^{*}b^{*} = 1 \quad (\text{T's } 3, 10). \end{cases}$$

This proves T21.

T21′. $(ab)^{*} = a^{*} \,\Delta\, b^{*}$.

T22. $(a^{\circ}b^{\circ})^{\circ} = (a' \,\Delta\, b')'$ $\quad [a \neq 0, \neq 1;\ b \neq 0, \neq 1;\ a' \,\Delta\, b' \neq 0]$.

Proof. By T's 21, 12

$$(a^{\circ}b^{\circ})^{*} = a^{\circ *} \,\Delta\, b^{\circ *} \qquad [a \neq 1, b \neq 1].$$

The theorem then follows at once from T's 10, 2, 3. Theorem 22, it is observed, is another self-dual theorem.

T23. $(a^{\circ}b^{\circ})^{\circ} = a(a^{\prime\circ}b)^{\circ}$ $\quad [a \neq 0, \neq 1;\ a^{\circ}b^{\circ} \neq 1]$.

Proof. For $b = 0$ (and $a \neq 0, \neq 1$) the theorem follows from T's 2, 3. Therefore, comparing the domains of T's 23 and 22, we have only to prove that

$$(1)\quad a(a'^{\circ}b)^{\circ} = (a' \Delta b')' \quad \text{for} \quad (a \neq 0, \neq 1;\ b \neq 0, \neq 1;\ a^{\circ}b^{\circ} \neq 1).$$

Under these restrictions, however, C′ gives

$$(2)\qquad (a' \Delta b')' = a''(a'^{\circ}b'')^{\circ\prime\prime},$$

which yields (1) and hence T23, by use of T3.

T23′. $(a' \Delta b')' = a \,\Delta\, (a^{\circ\prime} \Delta b)' \qquad [a \neq 0, \neq 1;\ a' \Delta b' \neq 0].$

4. Definition of $+$. Of several available definitions for field addition, $+$, we have chosen one which is commutative by inspection. Considerably more compact (but unsymmetric) formulas for $+$ are derived as theorems. The "singularities" of different such formulas for addition will in general be different.

Definition E2.

$$\begin{cases} a + b = a \Delta b \Delta a^{\circ}b^{\circ} & [a \neq 1, b \neq 1]. \\ a + 1 = 1 + a = a \Delta a'^{\circ} & [a \neq 0, \neq 1]. \\ 1 + 1 = \tau \\ 1 + 0 = 0 + 1 = 1. \end{cases}$$

We note (see 1) that the dual of the definition E2 furnishes a definition of $a \underset{\circ}{+} b$.

Definition E2′.

$$\begin{cases} a \underset{\circ}{+} b = ab(a' \Delta b') & [a \neq 0, b \neq 0]. \\ a \underset{\circ}{+} 0 = 0 + a = aa^{\circ\prime} & [a \neq 0, \neq 1]. \\ 0 \underset{\circ}{+} 0 = \mu. \\ 0 \underset{\circ}{+} 1 = 1 + 0 = 0. \end{cases}$$

It is quite immaterial whether we elect to show that $(F, +, \times)$, with the $+$ from E2, or that $(F, \underset{\circ}{+}, \Delta)$, with $\underset{\circ}{+}$ from E2′, is equivalent to the usual definition of field, since either follows from the other by the duality theorem (i. e., from the dual-pair nature of our postulates). We shall prove the former.

Theorems 24 to 27 are immediate consequences of E2, B, B′.

T24. *$a + b$ is a unique element of F for each a, b of F.*

T25. $a + b = b + a.$

T26. $a + b = 1,\ a \neq 1,\ b \neq 1 : \rightarrow a^{\circ}b^{\circ} = 1.$

Conversely, $a^{\circ}b^{\circ} = 1,\ a \neq 1,\ b \neq 1 : \rightarrow a + b = 1.$

T27. $a + b = 0,\ a \neq 1,\ b \neq 1 : \rightarrow a^{\circ} \Delta b^{\circ} = a^{\circ}b^{\circ}.$

Conversely, $a^{\circ} \Delta b^{\circ} = a^{\circ}b^{\circ},\ a \neq 1,\ b \neq 1 : \rightarrow a + b = 0.$

T28. $a + b = a \Delta a'^{\circ}b \qquad [a \neq 0, \neq 1].$

Proof. For $b = 1$ the theorem is true by E2, T2. For $a \neq 0, \neq 1$

$$(1) \qquad a \Delta a'^{\circ}b = a \Delta b \Delta b^{\circ}a'^{\circ *} = a \Delta b \Delta b^{\circ}a^{\circ},$$

by T's 20, 10, 2. The right of (1) is $a + b$ for $a \neq 0, \neq 1, b \neq 1$, which proves the theorem.

T29. $a + b = a(\mu a'b)^{*} \qquad [a \neq 0].$

Proof. *Case* 1, $a = 1$. Here, by T2, we must show that

$$(1) \qquad 1 + b = (\mu b)^{*}.$$

For $b = 0$ or for $b = 1$, (1) follows from T's 2, 10, 18, E2. For $b \neq 0, \neq 1$ write $\mu = b'b'^{\circ\prime}b''^{\circ}$ (by T's 4, 3), whence, using T10, (1) becomes

$$1 + b = (b^{*}b''^{\circ})^{*}$$

or, by T's 21′, 12, 10,

$$(2) \qquad 1 + b = b \Delta b'^{\circ}.$$

But (2) is true by E2, which proves case 1. There remains only to prove the theorem for $a \neq 0, \neq 1$. Putting $\mu = aa'^{\circ}a^{\circ\prime}$ and using T's 28, 10, 3 we have to show that

$$(3) \qquad a \Delta a'^{\circ}b = a(a'^{\circ}a^{\circ\prime}b)^{*}.$$

If $a'^{\circ}a^{\circ\prime}b = 1$ each side of (3) reduces to 0 (by T's 3, 10, 2). If $a'^{\circ}a^{\circ\prime}b \neq 1$ we may use C′ on the left of (3) and hence have to show that

$$(4) \qquad a(a^{\circ}a'^{\circ\prime}b')^{\circ\prime} = a(a'^{\circ}a^{\circ\prime}b)^{*}.$$

But (4) follows at once from T10, since the left side may be written $a(a^{\circ\prime}a'^{\circ}b)'^{\circ\prime}$. This proves T29.

T30. $a^{*}(ba^{*\prime})^{*} = b^{*}(ab^{*\prime})^{*} \qquad [a \neq 1, b \neq 1].$

Proof. From C′ and the commutativity of Δ we have

$$a(a^{\circ}b')^{\circ\prime} = b(b^{\circ}a')^{\circ\prime} \qquad [a \neq 0, \neq 1;\ b \neq 0, \neq 1;\ a^{\circ}b' \neq 1]$$

i. e.,
$$b'(a^{\circ\prime}b)^{\prime\circ\prime} = a'(b^{\circ\prime}a)^{\prime\circ\prime} \quad [\quad \text{“} \qquad \text{“} \qquad \text{“} \quad].$$

i. e., (1)
$$b'(a^{\circ\prime}b)^{*} = a'(b^{\circ\prime}a)^{*} \text{ for } [\quad \text{“} \qquad \text{“} \qquad \text{“} \quad].$$

The domain of (1) also permits $a^{\circ}b' = 1$ ($\rightleftarrows b^{\circ}a' = 1$), in which case (1) is true (each side reducing to 0) by T's 3, 10, 2. Hence we have proved the formula (1) for its domain; that is, putting $b' = r^{*}$, $a' = s^{*}$, we have proved

$$(2) \qquad r^{*}(sr^{*\prime})^{*} = s^{*}(rs^{*\prime})^{*} \quad \text{for} \quad (r \neq 0, \neq 1;\ s \neq 0, \neq 1).$$

The domain of (2) also permits $r = 0$ as well as $s = 0$. In each of these cases, however, (2) is at once true by T's 10, 2, 3, (each side reducing to s^{*}, respectively to r^{*}). Hence (2) is proved for its domain, which completes T30.

T31. $(a + b) + c = a + (b + c)$.

Proof. If any of the elements $a, b, c, = 0$ the theorem is true by E2. Hence we assume $a, b, c, \neq 0$. *Case* 1: $a + b \neq 0$, $b + c \neq 0$. We may here use T29 for all $+$'s, and have

$$(a + b) + c = b(\mu b'a)^{*}\{\mu b'(\mu b'a)^{*\prime}c\}^{*}$$

$$a + (b + c) = b(\mu b'c)^{*}\{\mu b'(\mu b'c)^{*\prime}a\}^{*}.$$

Hence we must establish that

$$(1) \qquad b(\mu b'a)^{*}\{\mu b'(\mu b'a)^{*\prime}c\}^{*} = b(\mu b'c)^{*}\{\mu b'(\mu b'c)^{*\prime}a\}^{*}.$$

Put $\mu b'a = r$, $\mu b'c = s$. Now $r \neq 1$, $s \neq 1$ since $r = 1 \rightarrow a + b = 0$ (T's 29, 10, 3), and similarly $s = 1 \rightarrow b + c = 0$, contrary to hypothesis. Therefore (1) is proved since, using T3, it is equivalent to

$$r^{*}(sr^{*\prime})^{*} = s^{*}(rs^{*\prime})^{*}$$

which is true by T30. This proves case 1. *Case* 2: $a + b = 0$, $b + c \neq 0$. (This is of course equivalent to the case: $a + b \neq 0$, $b + c = 0$). Now $a + b = 0 \rightarrow \mu ab' = 1$ (T's 29, 10). Therefore we must show that

$$0 + c = b(\mu b'c)^{*}\{\mu b'(\mu b'c)^{*\prime}a\}^{*} \quad \text{if} \quad \mu ab' = 1,$$

i. e. (by T's 6, 3, 2, E2) that

$$c = b(\mu b'c)^{*}(\mu b'c)^{*'*}. \tag{2}$$

Putting $b'c = h$, noticing that $\mu h \neq 0, \neq 1$ (by T's 29, 3, since $b + c \neq 0$), and using T's 3, 10, 17, (2) reduces to

$$h' = (\mu h)^{'\circ}(\mu h)^{\circ'}.$$

This however is evident, since by T4

$$\mu = \mu h(\mu h)^{'\circ}(\mu h)^{\circ'}.$$

This proves case 2. *Case* 3: $a + b = 0$, $b + c = 0$. Here, using T's 29, 10, we must prove that

$$\mu a'b = \mu c'b (= 1) \ : \rightarrow c = a.$$

This is immediate by T3, and completes the proof of T31.

T32. $0 + a \ (= a + 0) = a.$

Proof. E2, T's 2, 3.

T33. *$a + x = c$ has a solution.*

Proof. For $a = 0$, $x = c$ is a solution (T32). For $a \neq 0$,

$$x = a\mu(ca')^{*}$$

is a solution, by T's 29, 3, 6, 12.

T34. $a(b + c) = ab + ac.$

Proof. If a, b or $c = 0$ the theorem follows from T2 and T32. Assume, then, $a \neq 0$, $b \neq 0$, $c \neq 0$, which further implies $ab \neq 0$, $ac \neq 0$ (T3). Then by T29,

$$a(b + c) = ab(\mu b'c)^{*} \tag{1}$$

and

$$ab + ac = ab\{\mu(ab)'ac\}^{*}, = ab(\mu b'c)^{*}. \tag{2}$$

The theorem follows by comparing (1) and (2).

From B and T's 2, 24, 25, 31, 32, 33, 34 one then has

T35. *With $+$ defined by E2, $(F, +, \times)$ is a field with* 1 *as its unit and* 0 *as its zero element.*

5. Equivalence. Theorem 35 together with the observation immediately following the postulates establishes the equivalence, in the sense of interdefinability, of the classical $(F, +, \times; 0, 1)$ and the dual-symmetric $(F, \Delta, \times; 0, 1)$ characterization of the concept field via the connecting definitions

$$\begin{array}{rcll}
\multicolumn{3}{c}{(F, +, \times; 0, 1) \parallel (F, \Delta, \times; 0, 1)} & \\
F \leftarrow & \text{defined by} & \rightarrow F & \\
0 \leftarrow & \text{``} & \rightarrow 0 & \\
1 \leftarrow & \text{``} & \rightarrow 1 & \\
a \times b \leftarrow & \text{``} & \rightarrow a \times b & \\
a + b - ab \leftarrow & \text{``} & \leftarrow a \Delta b & \\
a + b \rightarrow & \text{``} & \rightarrow a \Delta b \Delta a^\circ b^\circ & (\text{for } a \neq 1, b \neq 1) \\
a + 1, = 1 + a \rightarrow & \text{``} & \rightarrow a \Delta a'^\circ & (\text{`` } a \neq 0, \neq 1) \\
1 + 1 \rightarrow & \text{``} & \rightarrow h \Delta h'^\circ \Delta h^{\circ\prime} & (\text{`` } h \neq 0, \neq 1) \\
1 + 0, = 0 + 1 \rightarrow & \text{``} & \rightarrow 1. &
\end{array}$$

But more than this is true,—the two characterizations are *strongly*[4] equivalent, that is, also the (above) *definitions* used in each system (to establish the equivalence) are provable as theorems of the other system. Only two such theorems come into question here, namely

THEOREM EQ 1. $a + b - ab = a \Delta b$,
a theorem of our dual-symmetric characterization $(F, \Delta, \times)$, and

THEOREM EQ 2.

$$\begin{array}{rl}
a + b = a \Delta b \Delta a^\circ b^\circ & (\textit{for } a \neq 1, b \neq 1) \\
a + 1 = 1 + a = a \Delta a'^\circ & (\text{`` } a \neq 0, \neq 1) \\
1 + 1 = h \Delta h'^\circ \Delta h^{\circ\prime} & (\text{`` } h \neq 0, \neq 1) \\
0 + 1 = 1 + 0 = 1, &
\end{array}$$

a theorem of the classical $(F, +, \times)$ characterization of field. The latter is immediately verified upon substituting $a + b - ab$, $a/a - 1$ and $1/a$ for $a \Delta b$, a° and a' respectively. We prove Theorem Eq 1, or, what is clearly the same, the formula

$$(1) \qquad a + b = (a \Delta b) + ab.$$

Proof. For each of the cases $a = 0$, $a = 1$, $b = 0$, $b = 1$, (1) is true by E2, T2. Assume, then, $a \neq 0, \neq 1$; $b \neq 0, \neq 1$. *Case* 1: $a \Delta b \neq 0$. In this case, by T28,

---

[4] The explicit recognition of the important distinction between mere "interdefined" and "strong" equivalence of postulates (and related observations on equivalence) is due to A. Tarski in a paper not yet published.

$$(a \Delta b) + ab = a \Delta b \Delta (a \Delta b)^{\prime\circ} ab \tag{2}$$

$$a + b = a \Delta b \Delta a^{\circ} b^{\circ}. \tag{3}$$

Hence, simplifying by T3, we must prove that

$$(a \Delta b)^{\prime\circ} ab = a^{\circ} b^{\circ}$$

or, using C′, that

$$\{a^{\prime}(a^{\circ} b^{\prime})^{\circ}\}^{\circ} ab = a^{\circ} b^{\circ}. \tag{4}$$

Now $a^{\prime}(a^{\circ} b^{\prime})^{\circ} \neq 1$, since by T3

$$a^{\prime}(a^{\circ} b^{\prime})^{\circ} = 1 \rightarrow a^{\circ} b^{\prime} = a^{\circ}, \rightarrow b = 1,$$

contrary to hypothesis. Therefore we may use T23, and (4) becomes

$$a^{\circ} b^{\prime} \{(a^{\circ\prime} b)^{\circ} a^{\prime\circ}\}^{\circ} ab = a^{\circ} b^{\circ}$$

i. e., by T3,

$$a\{(a^{\circ\prime} b)^{\circ} a^{\prime\circ}\}^{\circ} = b^{\circ}. \tag{5}$$

Again $(a^{\circ\prime} b)^{\circ} a^{\prime\circ} \neq 1$, since

$$(a^{\circ\prime} b)^{\circ} a^{\prime\circ} = 1 \rightarrow a^{\circ\prime} b = a^{\prime\circ\prime\circ} = a^{\circ\prime} \text{ (by T's 9, 3)}, \rightarrow b = 1,$$

contrary to hypothesis. So we may again apply T23, and (5) becomes

$$a a^{\prime}(a^{\circ} a^{\circ\prime} b)^{\circ} = b^{\circ}, \tag{6}$$

which is obviously true, by T3. This proves case 1. *Case* 2: $a \Delta b = 0$, (equivalent to $b = a^{\circ}$). Using E2 and T32 we must here show that

$$a a^{\circ} = a \Delta a^{\circ} \Delta a a^{\circ}. \tag{7}$$

But (7) is obvious by T3. This completes Theorem Eq 1, and establishes the strong-equivalence of the ordinary $(F, +, \times)$ and the dual-symmetric $(F, \Delta, \times)$ characterizations of the concept field.

6. Set II of postulates. The primitive notions of set II are $(F, \Delta, \times)$, where Δ, $\times$ as before, correspond (in realization) to the two dual field multiplications. The field complement, *, as a definition, is here involved in the formulation of certain of the postulates, and this requires that the dual of a proposition

$$P(1, 0;\ \times, \Delta, {}^{\circ}, {}^{\prime}, {}^{*})$$

be *defined* as

$$P(0, 1;\ \Delta, \times, {}^{\prime}, {}^{\circ}, {}^{*}),$$

where, it is to be noted, the invariance of * (unlike set I) is here a part of the *definition* of $dl\,P$. The postulates of set II are then seen to occur in dual pairs, (with N and P7 self-dual), and hence, as in I, it is only necessary to prove one of a pair of dual theorems. References are given in full with the exception of those to N, P1, P1′, which are omitted. The postulates are:

N. *F contains at least three elements.*

P1. *ab is a unique element of F whenever a, b are in F.*

P1′. *$a \Delta b$ is a unique element of F whenever a, b are in F.*

P2. $a(bc) = (ab)c$.

P2′. $a \Delta (b \Delta c) = (a \Delta b) \Delta c$.

P3. $ab = ba$.

P3′. $a \Delta b = b \Delta a$.

P4. *There is an element 1 in F such that $a1 = a$ for every a in F.*

P4′. *There is an element 0 in F such that $a \Delta 0 = a$ for every a in F.*

P5. *For each $a \neq 0$ in F there is an a' in F such that $aa' = 1$.*

P5′. *For each $a \neq 1$ in F there is an a° in F such that $a \Delta a^\circ = 0$.*

Definition D1.

$$\begin{cases} a^* = a^{\circ\prime\circ} & [a \neq 0, \neq 1]. \\ 1^* = 0. \\ 0^* = 1. \end{cases}$$

P6. $ab = a \Delta a^\circ b^* \qquad [a \neq 1]$.

P6′. $a \Delta b = a(a' \Delta b^*) \qquad [a \neq 0]$.

P7. $(a^\circ a' a^*)^\circ = (b' \Delta b^\circ \Delta b^*)' \qquad [a \neq 0, \neq 1;\ b \neq 0, \neq 1]$.

T0. (THEOREM 0). $0 \neq 1$.

Proof. (1) $a0 = 0$, since: if $a \neq 1$, then

$$a0 = a \Delta a^\circ 0^* = a \Delta a^\circ 1 = a \Delta a^\circ = 0,$$

by P6, D1, P4, P5′: and if $a = 1$, then $01 = 0$, by P4, P3. If $0 = 1$, we have the contradiction: for $a \neq 0$, $a = a1 = a0 = 0$, by N. P4, hypothesis, and (1).

T1. *$(F, \times)$ without 0 is an Abelian group with 1 as its unit element.*

Proof. T0, P's 1, 2, 3, 4, 5.

T1′. *(F, Δ) without 1 is an Abelian group with 0 as its unit element.*

T2. $0a = a0 = 0$.

Proof. P3 and (1) of T0.

T2′. $1 \Delta a = a \Delta 1 = 1$.

T3. $a^{**} = a$.

Proof. If $a \neq 0, \neq 1$, then $a^{**} = a^{\circ\prime\circ\circ\prime\circ} = a^{\circ\prime\prime\circ} = a^{\circ\circ} = a$, by D1, T's 1′, 1, 1′. If $a = 0$, then $0^{**} = 1^{*} = 0$, by D's 1, 1. If $a = 1$, then $1^{**} = 0^{*} = 1$, by D's 1, 1.

T4. $(ab)^{*} = a^{*} \Delta b^{*}$.

Proof. If $a \neq 0; \neq 1$, then $a^{\circ}b^{*} = a^{\circ} \Delta ab$; by P6, T1′; hence

$$b^{*} = a^{\circ\prime}(a^{\circ} \Delta ab) = a^{\circ\prime}(a^{\circ\prime\prime} \Delta ab) = a^{\circ\prime}\{a^{\circ\prime\prime} \Delta (ab)^{**}\} = a^{\circ\prime} \Delta (ab)^{*}$$

by T's 1, 1, 3, P6′; hence, $(ab)^{*} = a^{\circ\prime\circ} \Delta b^{*} = a^{*} \Delta b^{*}$, by T1′, D1. If $a = 1$, then $(1b)^{*} = b^{*} = 0 \Delta b^{*} = 1^{*} \Delta b^{*}$, by T's 1, 1′, D1. If $a = 0$, then $(0b)^{*} = 0^{*} = 1 = 1 \Delta b^{*} = 0^{*} \Delta b^{*}$, by T2, D1, T2′, D1.

T4′. $(a \Delta b)^{*} = a^{*}b^{*}$.

T5. $a^{\circ}a^{\prime *} = 1 \qquad [a \neq 0, \neq 1]$.

Proof. $1 = aa^{\prime} = a \Delta a^{\circ}a^{\prime *}$, by T1, P6; hence, $a^{\circ}a^{\prime *} = 1 \Delta a^{\circ} = 1$, by T's 1′, 2′.

T5′. $a^{\prime} \Delta a^{\circ *} = 0, \qquad [a \neq 0, \neq 1]$.

T6. $a^{\prime *} = a^{\circ\prime} \qquad [a \neq 0, \neq 1]$.

Proof. T's 5, 1.

T6′. $a^{\circ *} = a^{\prime\circ} \qquad [a \neq 0, \neq 1]$.

T7. $a^{*} = a^{\prime\circ\prime} = a^{\circ\prime\circ} \qquad [a \neq 0, \neq 1]$.

Proof. $a^{\circ\prime\circ} = a^{*} = a^{\prime\prime *} = a^{\prime\circ\prime}$, by D1, T's 1, 6.

T8. $a(a^{\prime}b^{*})^{*} = b(b^{\prime}a^{*})^{*} \qquad [a \neq 0,\ b \neq 0]$.

Proof.

$$\begin{aligned} a(a'b^*)^* &= a(a'^* \Delta b^{**}) = a(a'^* \Delta b) = a(b \Delta a'^*) \\ &= a\{b(b' \Delta a'^{**})\} = a\{b(b' \Delta a')\} = b\{a(b' \Delta a')\} \\ &= b\{a(a' \Delta b')\} = b\{a(a' \Delta b'^{**})\} = b(a \Delta b'^*) \\ &= b(a^{**} \Delta b'^*) = b(a^*b')^* = b(b'a^*)^*, \end{aligned}$$

by T's 4, 3, 1′, P6′, T's 3, 1, 1′, 3, P6′, T's 3, 4, 1.

T8′. $a \Delta (a^\circ \Delta b^*)^* = b \Delta (b^\circ \Delta a^*)^*$ $[a \neq 1, b \neq 1]$.

T9. $a^\circ a' a^* =$ *a constant* $[a \neq 0, \neq 1]$.

Proof. P7, T's 1, 1′.

T9′. $a' \Delta a^\circ \Delta a^* =$ *a constant* $[a \neq 0, \neq 1]$.

T10. *If* $a \neq 0, \neq 1$ *then:* $a^\circ \neq 0, \neq 1$; $a' \neq 0, \neq 1$; $a^* \neq 0, \neq 1$.

Proof. T's 1′, 1′, 1, 1′, 3, D1.

Definition D2. $\mu = a^\circ a' a^*$ $[a \neq 0, \neq 1]$.

Definition D2′. $\tau = a' \Delta a^\circ \Delta a^*$ $[a \neq 0, \neq 1]$.

T11. $\mu \neq 0$.

Proof. D2, T's 10, 1.

T11′. $\tau \neq 1$.

T12. (i) $\mu = a^\circ a' a^* = a a^{\circ\prime} a^{\prime\circ}$; $\tau = a' \Delta a^\circ \Delta a^* = a \Delta a^{\prime\circ} \Delta a^{\circ\prime}$ $[a \neq 0, \neq 1]$.

(ii) $\mu' = \mu$; $\tau^\circ = \tau$.

(iii) $\mu^\circ = \tau'$ $[\mu \neq 1, (\rightleftarrows \tau \neq 0)]$.

(iv) $\mu^* = \tau$; $\tau^* = \mu$.

(v) $\mu\mu = 1$; $\tau \Delta \tau = 0$.

(vi) $\mu a = a^* a^\circ$ $[a \neq 1]$; $\tau \Delta a = a^* \Delta a'$ $[a \neq 0]$.

(vii) $\mu a' = a'^* a^{\circ *}$ $[a \neq 0, \neq 1]$; $\tau \Delta a^\circ = a^{\circ *} \Delta a'^*$ $[a \neq 0, \neq 1]$.

Proof. Only the first parts require proof, since the second parts are their duals. (i): by D2 (a° for a) and T6′. (ii): by i, T's 1, 7. (iii): by P6, D2, D2′. (iv): by D1, iii, T1, ii. (v) by ii, T1. (vi): by i, T1. (vii): by vi, T6′.

T13. $a(\mu a')^* = (\mu a)^*$ $[a \neq 0]$.

Proof.

$$a(\mu a')^* = a(a'^*a^{\circ *})^* = a(a'^{**} \Delta a^{\circ **}) = a(a' \Delta a^{\circ **})$$
$$= a \Delta a^{\circ *} = a^{**} \Delta a^{\circ *} = (a^*a^\circ)^* = (\mu a)^*,$$

by T12 (vii), T's 4, 3, P6′, T's 3, 4, 12 (vi).

T13′. $a \Delta (\tau \Delta a^\circ)^* = (\tau \Delta a)^* \qquad [a \neq 1]$.

T14. $a\{\mu a'(\mu b)^*\}^* = b\{\mu b'(\mu a)^*\}^* \qquad [a \neq 0, b \neq 0]$.

Proof. $(\mu a)\{(\mu a)'(\mu b)^*\}^* = (\mu b)\{(\mu b)'(\mu a)^*\}^*$, by T's 8, 11. Hence

$$\mu a\{\mu' a'(\mu b)^*\}^* = \mu b\{\mu' b'(\mu a)^*\}^*,$$

by T1; hence the theorem, by T's 1, 11, 12 (ii), 1.

T14′. $a \Delta \{\tau \Delta a^\circ \Delta (\tau \Delta b)^*\}^* = b \Delta \{\tau \Delta b^\circ \Delta (\tau \Delta a)^*\}^* \; [a \neq 1, b \neq 1]$.

Definition D3. $\begin{cases} a + b = a(\mu a'b)^* \qquad [a \neq 0]. \\ 0 + a = a. \end{cases}$

T15. $a + 1 = 1 + a$.

Proof. If $a \neq 0$, then

$$a + 1 = a(\mu a'1)^* = a(\mu a')^* = (\mu a)^* = 1(\mu a'a)^* = 1 + a,$$

by D3, T's 1, 13, 1, D3. If $a = 0$, then

$$1 + 0 = 1(\mu 1'0)^* = (\mu 0)^* = 0^* = 1 = 0 + 1,$$

by D3, T's 1, 2, D1, D3.

T16. *$a + b$ is a unique element of F for all a, b of F.*

Proof. D3, T1.

T17. $a + 0 = 0 + a = a$.

Proof. If $a \neq 0$, then

$$a + 0 = a(\mu a'0)^* = a0^* = a1 = a = 0 + a,$$

by D3, T2, D1, T1, D3. If $a = 0$, then $0 + 0 = 0$, by D3.

T18. $a(b + c) = ab + ac$.

Proof. If $a \neq 0$, $b \neq 0$, then

$$ab + ac = ab\{\mu(ab)'ac\}^* = ab(\mu b'c)^* = a\{b(\mu b'c)^*\} = a(b + c),$$

by D3, T's 1, 1, D3. If $a = 0$, $b \neq 0$, then

$$0(b + c) = 0 = 0 + 0 = 0b + 0c,$$

by T2, D3, T2. If $a \neq 0$, $b = 0$, then

$$a(0 + c) = ac = 0 + ac = a0 + ac,$$

by D's 3, 3, T2. If $a = 0$, $b = 0$, then

$$0(0 + 0) = 0 = 0 + 0 = 00 + 00,$$

by T2, D3, T2.

T19. $a + b = b + a$.

Proof. If $a \neq 0$, then

$$a + b = a1 + aa'b = a(1 + a'b) = a(a'b + 1) = aa'b + a1 = b + a,$$

by T's 1, 18, 15, 18, 1. If $a = 0$, the theorem is immediate, by T17.

T20. $a + (1 + b) = b + (1 + a)$.

Proof. If $a \neq 0$, $b \neq 0$, then

$$\begin{aligned} a + (1 + b) &= a + 1(\mu 1'b)^* = a + (\mu b)^* = a\{\mu a'(\mu b)^*\}^* \\ &= b\{\mu b'(\mu a)^*\}^* = b\{\mu b'1(\mu 1'a)^*\}^* = b + (1 + a), \end{aligned}$$

by D3, T1, D3, T's 14, 1, D3. If $a \neq 0$, $b = 0$, then

$$a + (1 + 0) = a + 1 = 1 + a = 0 + (1 + a),$$

by T's 17, 15, 17. If $a = 0$, $b \neq 0$, then

$$0 + (1 + b) = 1 + b = b + 1 = b + (1 + 0),$$

by T's 17, 15, 17. If $a = 0$, $b = 0$, then

$$0 + (1 + 0) = 0 + (1 + 0).$$

T21. $a + (b + c) = c + (b + a)$.

Proof. If $b \neq 0$, then

$$\begin{aligned} a + (b + c) &= bab' + (b + bb'c) = b\{ab' + (1 + b'c)\} \\ &= b\{b'c + (1 + ab')\} = bb'c + (b + bab') = c + (b + a), \end{aligned}$$

by T's 1, 18, 20, 18, 1. If $b = 0$, then

$$a + (0 + c) = a + c = c + a = c + (0 + a),$$

by T's 17, 19, 17.

T22. $a + (b + c) = (a + b) + c$.

Proof. T's 21, 19.

T23. *For each a there is an $\bar{a}$ such that $a + \bar{a} = 0$.*

Proof. If $a \neq 0$, then $\bar{a} = \mu a$. For,

$$a + \mu a = a\{\mu a'(\mu a)\}^* = a(\mu\mu)^* = a1^* = a0 = 0,$$

by D3, T's 1, 12(v), D1, T2. If $a = 0$, then $\bar{a} = 0$, by T17.

From T's 16, 17, 18, 19, 22, 23, N, P's 1, 2, 3, 4, 5 we have

T24. *With $+$ defined by D3, $(F, +, \times)$ is a field with 1 as its unit, and 0 as its zero element.*

With the help of P6 and T's 3, 4, 7, Postulate C of set I can readily be derived, and the equivalence of sets I and II established.

UNIVERSITY OF CALIFORNIA,
BERKELEY.

A MATHEMATICAL THEORY OF DILATANCY.*

By M. Reiner.

Dilatancy is "that property of a material by which it increases in volume when it is stirred" or, more specifically, when it is sheared. Thus "in walking over wet sands the gloss disappears and the sand appears to dry out wherever the foot is placed." This is due to the increase in volume of the bulk of the sand, i. e., the increase of the *volume of the voids* into which the water is drawn. Dilatancy was first scientifically described by Reynolds in 1885, who coined the term. In the following a mathematical theory of dilatancy is expounded from the phenomenological point of view, i. e., no account will be taken of the internal structure of the material to which the phenomenon of dilatancy may be due. It will be shown that on purely mathematical grounds dilatancy should be present in a generalized Newtonian liquid.

1. A *Newtonian liquid* is a material which is perfectly elastic under a hydrostatic pressure (or tension) and perfectly viscous in distortion. By the latter is meant that all strainwork in distortion is entirely dissipated. This distinguishes the Newtonian liquid from that of Maxwell which has (relaxing) elasticity of shape. In a "*simple*" Newtonian liquid "*the resistance which arises from the lack of slipperiness of the parts of the liquid, other things being equal, is proportional to the velocity with which the parts of the liquid are separated from one another*" (Newton, *Principia*, Lib. ii, Sect. ix). In the "*generalized*" Newtonian liquid we drop the linearity and substitute for it any functional relationship between the deviators of stress and velocity-strain.[1] The generalized Newtonian liquid is accordingly one case of "non-linear" rheological bodies, the other important one being the non-linear or generalized Hookean, as we may say, elastic solid. The theory of the latter was treated by Hamel and brought to a conclusion by Murnaghan. It is bound up with the theory of finite strain. We define the strain tensor by $\epsilon_{xx} = \partial u/\partial x - \frac{1}{2}\{(\partial u/\partial x)^2 + (\partial v/\partial y)^2 + (\partial w/\partial z)^2\}$ etc. (i. e., the usual Cauchy definition, where u, v, w[2] are the components of the displacement vector) which, for small relative displacements, become $\partial u/\partial x$ etc.; and we define the strain of

---

* Received August 6, 1944.

[1] By "velocity-strain" we understand the space derivative of the velocity-vector, in contradistinction to the ordinary or *displacement*-strain.

[2] We shall use in the present paper Murnaghan's notation.

the velocity vector $(\dot{u}, \dot{v}, \dot{w})$ by $e_{xx} = \partial\dot{u}/\partial x$ etc., the higher terms being absent because the displacements $\dot{u}dt$ etc. in this case are infinitesimal; then evidently e_{xx} is generally not equal to $\dot{\epsilon}_{xx}$. This is expressed by Murnaghan in the words "the space derivative of the virtual displacement vector . . . is only in the infinitesimal theory (equal to) . . . the variation of the strain tensor." Hencky has shown how this difficulty can be overcome in the case of fixed principal axes of strain by a different definition of the strain, which for infinitesimal strain approximates the Cauchy strain. However, fortunately, we need not enter into these difficulties here. *The relative displacements upon which the stresses in a viscous liquid* (apart from the hydrostatic pressure) *depend are always infinitesimal by definition and* the theory is thereby considerably simplified.

2. The problem stated. In order to take account of the fundamentally different properties of a Newtonian liquid with respect to isotropic and deviatoric stress and strain components, we resolve the tensors of stress $T^*{}_s{}^r$, strain $\epsilon_s{}^r$ and velocity-strain $e_s{}^r$ as follows:

(1) $$T^*{}_s{}^r = T^*\delta_s{}^r + T'{}_s{}^r,$$

(2) $$\epsilon_s{}^r = \epsilon\delta_s{}^r + \epsilon'{}_s{}^r,$$

(3) $$e_s{}^r = e\delta_s{}^r + e'{}_s{}^r.$$

This resolution is defined by

(4) $$T'{}_a{}^a = \epsilon'{}_a{}^a = e'{}_a{}^a \equiv 0,$$

while the scalars $3T^*$, 3ϵ and $3e$ are the first invariant of the tensors $T^*{}_s{}^r$, $\epsilon_s{}^r$ and $e_s{}^r$, respectively, or

(5) $$T^*{}_a{}^a = 3T^*, \qquad \epsilon_a{}^a = 3\epsilon, \qquad e_a{}^a = 3e.$$

As the Newtonian liquid possesses elasticity of volume, but not elasticity of form, T^* contains a part ($-p$, where p is called the hydrostatic pressure) which is a function of ϵ

(6) $$p = F(\epsilon)$$

while $\epsilon'{}_s{}^r$ does not enter into the expression of $T^*{}_s{}^r$ in terms of strain. The hydrostatic pressure is not identical with the isotropic component of the stress tensor, but, as its name implies, only such part of it as remains for $e_s{}^r = 0$. (6) can also be written in the form

(6′) $$p = -3\kappa\epsilon$$

where κ is the elastic bulk modulus and is itself a function of ϵ. We, therefore, write

$$(7) \qquad T^{*}{}_{s}{}^{r} = (T - p)\delta_s{}^r + T'{}_s{}^r = (T + 3\kappa\epsilon)\delta_s{}^r + T'{}_s{}^r,$$

or

$$(8) \qquad T^{*}{}_{s}{}^{r} - 3\kappa\epsilon\delta_s{}^r = T_s{}^r = T\delta_s{}^r + T'{}_s{}^r,$$

but we shall not deal in the present paper with (6). This is common to both solids and liquids and has received complete treatment by Murnaghan and Birch. Accordingly we write $T_s{}^r$ for $T^{*}{}_{s}{}^{r} - 3\kappa\epsilon\delta_s{}^r$ and consider in the following $T_s{}^r$ only.[3] The rheological equation (compare Hersey) of the Newtonian liquid in its most general form can now be written as follows:

$$(9) \qquad T_s{}^r = R(\epsilon, e_s{}^r)$$

or, in view of (6),

$$(10) \qquad T_s{}^r = f(p, e_s{}^r).$$

The elaboration of (10) can be carried out in two stages. In the first stage we write

$$(11) \qquad T_s{}^r = f(e_s{}^r),$$

which means that the constants appearing in this equation will generally be functions of the hydrostatic pressure or of the density of the material. The dependence of the coefficient of viscosity upon the hydrostatic pressure is well known. However, just as in the classical hydrodynamics of viscous liquids the coefficient of viscosity is treated as a constant and its dependence upon p relegated to a secondary consideration, in the same way we shall in the present paper limit ourselves to (11). Our problem is to develop the function f in accordance with certain postulates to which we are led by physical observations.

3. The Murnaghan method. We first follow the way taken by Hamel and Murnaghan in accordance with which the rheological equation is derived from the two thermodynamical laws. For isothermal processes both are combined in the Gibbs-Helmholtz equation

$$(12) \qquad \begin{aligned} \delta w &= \rho\delta\Phi + \rho\delta\psi \\ \delta\psi &\geq 0 \end{aligned}$$

where w is the strainwork per unit volume, Φ is the intrinsic[4] free energy-density and ψ is the bound energy-density. The second part of (12) expresses the condition that while in the case of thermodynamic equilibrium ψ vanishes, in other cases it can only increase (compare Weissenberg). Generally, in a rheological process of every *real* material both functions Φ and ψ will be involved, but two *ideal* materials can be conceived. For the first ψ vanishes identically. This is the elastic solid, which has been dealt with by Murnaghan. For the second Φ vanishes in distortion or, more generally, in the absence of p. This is the viscous liquid with which we are dealing in the present paper. In the latter case we can introduce into (12) $\delta\Phi = 0$, when we get $\delta w = \rho\delta\psi$ and therefore

---

[3] The tensor $T_s{}^r$ is the viscous resistance of the liquid in its most general sense.

[4] I. e., after deduction of the kinetic energy produced. This is effected in viscometry by means of the so-called kinetic energy correction.

$$\delta w \geq 0. \tag{12'}$$

There is no complete analogy between Φ and ψ. While the former can be expressed as a function of mechanical variables with time not appearing explicitly, or $\partial\Phi/\partial t = 0$, this is generally not the case with regard to ψ. Nevertheless, as Rayleigh has shown, there exists a function $\chi = \chi(\dot{\psi})$ in the case of a simple Newtonian liquid, such that the hydrodynamic stress tensor can be expressed as a derivative of χ in analogy to the elastic stress tensor. In particular he found for such liquids

$$\chi = \rho\dot{\psi}/2 = w/2. \tag{13}$$

He called the function χ the "dissipation function." When such a function exists, a theory of generalized Newtonian liquids in analogy to Murnaghan's theory of elastic solids is possible. *In the present section we shall postulate the existence of a dissipation function* and see where this leads us, while in the next section we shall drop this postulate and proceed by a more general method.

We therefore write in analogy to Murnaghan, but keeping in mind the simplification introduced by the fact that in our case the relative displacements are infinitesimal,

$$T_s{}^r = \partial\chi/\partial e_s{}^r. \tag{14}$$

The dissipation function is, in analogy to Φ, a function

$$\chi = \chi(I_1, I_2, I_3) \tag{15}$$

of the invariants of the tensor $e_s{}^r$, while the temperature does not appear explicitly because rheological processes are assumed to be isothermal.

From (14) we find in analogy to Murnaghan

$$T_s{}^r = \delta_s{}^r\partial\chi/\partial I_1 + (\delta_s{}^r I_1 - e_s{}^r)(\partial\chi/\partial I_2) + \zeta_s{}^r I_3 \partial\chi/\partial I_3. \tag{16}$$

where the tensor $\zeta_s{}^r$ is defined by

$$\zeta_a{}^r e_s{}^a = \delta_s{}^r. \tag{17}$$

In order to apply (11) to the generalized Newtonian liquid, we calculate the strain-power per unit volume, which we denote by $\dot{w}$. This is equal to the trace of the inner product of the stress- and the velocity-strain tensor, or

$$\dot{w} = T_\beta{}^\alpha e_\alpha{}^\beta. \tag{18}$$

Making use of the relations

$$\begin{aligned} \delta_\beta{}^\alpha e_\alpha{}^\beta &= e_\alpha{}^\alpha = I_1 \\ e_\beta{}^\alpha e_\alpha{}^\beta &= I_1{}^2 - 2I_2 \\ \zeta_\beta{}^\alpha e_\alpha{}^\beta &= \delta_\alpha{}^\alpha = 3 \end{aligned} \tag{19}$$

we find from (18) and (16)

$$\dot{w} = I_1 \partial\chi/\partial I_1 + 2I_2 \partial\chi/\partial I_2 + 3I_3 \partial\chi/\partial I_3. \tag{20}$$

We now carry out the resolution of both tensors $T_s{}^r$ and $e_s{}^r$ in accordance with (1) and (3). This carries with it a resolution of the strain-power in accordance with the following equation,

$$T_\beta{}^\alpha e_\alpha{}^\beta = (T\delta_\beta{}^\alpha + T'_\beta{}^\alpha)(e\delta_\alpha{}^\beta + e'_\alpha{}^\beta) = 3Te + T'_\beta{}^\alpha e'_\alpha{}^\beta \tag{21}$$

into a volumetric component

$$\dot{w}_V = 3Te \tag{22}$$

and a deviatoric component

$$\dot{w}' = T'_\beta{}^\alpha e'_\alpha{}^\beta. \tag{23}$$

Considering the stress first, we find from (16) by means of

$$\delta_\alpha{}^\alpha = 3, \qquad e_\alpha{}^\alpha = I_1, \qquad \zeta_\alpha{}^\alpha I_3 = I_2, \tag{24}$$

$$T = T_\alpha{}^\alpha/3 = \partial\chi/\partial I_1 + \tfrac{2}{3} I_1 \partial\chi/\partial I_2 + I_2/3 \cdot (\partial\chi/\partial I_3) \tag{25}$$

and therefore

$$T'_s{}^r = \delta_s{}^r/3 \cdot (I_1 \partial\chi/\partial I_2 - I_2 \partial\chi/\partial I_3) - e_s{}^r \partial\chi/\partial I_2 + \zeta_s{}^r I_3 \partial\chi/\partial I_3. \tag{26}$$

Then we introduce the resolution of the strain. This affects the strain-invariants I_1, I_2 and I_3. They are replaced by

(27)[5]

$$\begin{aligned} e_1 &= \tfrac{1}{3} I_1 \\ I'_2 &= I_2 - 3e^2 \\ I'_3 &= I_3 - I_2 e + 2e^3 \end{aligned}$$

From the transformation equation

$$\partial\chi/\partial I_r = \partial\chi/\partial I'_a \cdot \partial I'_a/\partial I_r, \tag{28}$$

we get

---

[5] Note that I'_i is not the same as Murnaghan's I'_i.

$$
\begin{aligned}
&\partial\chi/\partial I_1 = \tfrac{1}{3}\partial\chi/\partial e - 2e\partial\chi/\partial I'_2 - \tfrac{1}{3}(I'_2 - 3e^2)\partial\chi/\partial I'_3 \\
(29)\qquad &\partial\chi/\partial I_2 = \partial\chi/\partial I'_2 - e\partial\chi/\partial I'_3 \\
&\partial\chi/\partial I_3 = \partial\chi/\partial I'_3
\end{aligned}
$$

and our equations (25) and (26) become

$$
\begin{aligned}
&T = \tfrac{1}{3}\partial\chi/\partial e \\
(31)\qquad &T'{}_s{}^r = -\delta_s{}^r[(I'_2 + 3e^2)/3 \cdot \partial\chi/\partial I'_3] - e'{}_s{}^r[\partial\chi/\partial I'_2 - e\partial\chi/\partial I'_3] \\
&\qquad\qquad + \zeta_s{}^r[(I'_3 + I'_2 e + e^3)\partial\chi/\partial I'_3],
\end{aligned}
$$

while. $\dot{w}_V$ and $\dot{w}'$ are from (22) and (23)

$$(32)\qquad \dot{w}_V = e\partial\chi/\partial e$$

$$(33)\qquad \dot{w}' = 2I'_2\partial\chi/\partial I'_2 + 3I'_3\partial\chi/\partial I'_3.$$

Murnaghan's next step is to consider ϕ as a function $\phi(\epsilon_s{}^r)$ of the strain components through the invariants and to expand it into a power series, neglecting terms of orders of magnitude greater than an agreed upon order. Elastic strains of even very deformable materials are small in comparison with the velocity-strains of viscous liquids which find their limit in turbulence only. A power-series development will therefore not carry us far. We should, however, naturally expect the lowest approximation of the generalized Newtonian liquid to yield the expressions for the simple Newtonian liquid. If we put

$$(34)\qquad \chi = ae + be^2 + cI'_2 + \cdots$$

we find

$$
\begin{aligned}
(35)\qquad &T = a/3 + 2be/3 + \cdots; \qquad T'{}_s{}^r = -ce'{}_s{}^r + \cdots; \\
&\dot{w} = ae + 2be^2 + 2cI'_2 + \cdots.
\end{aligned}
$$

For the case of rest this would give $T = a/3$, i. e., a hydrostatic stress, the consideration of which we have excluded. We therefore put $a = 0$ which gives $\chi = \dot{w}/2$ in accordance with Rayleigh's expression for the dissipation function (compare (13)). This, evidently, has a wider range than the classical viscous liquid, because for the latter, in order to get Stokes' expression for the stress, we have, in addition, to put $b = 0$ which gives $T = 0$ and $T_s{}^r = 2\eta(e_s{}^r - e\delta_s{}^r)$ with η, the coefficient of viscosity, $= -c/2$.

4. Application of a tensor analysis method.[6] The results of the previous chapter are valid only in the restricted case when there exists a dissipation function. In the present chapter we drop this restriction and proceed by pure tensor analysis of (11). Starting from this equation we note that on the left side there stands a mixed tensor of rank two. Then what we may call with Eddington the *principle of generalized dimensions* requires that all terms on the right side shall also be mixed tensors of rank two. The right side can therefore consist only of sums of mixed tensors of rank two multiplied by scalars and of inner products of such tensors which again are reduced to tensors of rank two. The general term of a development of the function f will therefore be of the form $e_\alpha{}^r e_\beta{}^\alpha e_\gamma{}^\beta \cdots e_s{}^\lambda \cdot f(I)$, where $f(I)$ is a function of the three invariants; and we can therefore write

---

[6] This treatment was suggested to me by Prof. Giulio Racah of the Hebrew University, Jerusalem.

$$(36) \quad T_s{}^r = f_0(I)\delta_s{}^r + f_1(I)e_s{}^r + f_2(I)e_\alpha{}^r e_s{}^\alpha + f_3(I)e_\alpha{}^r e_\beta{}^\alpha e_s{}^\beta + \cdots.$$

This would mean an infinite number of such terms. However, in view of the Cayley-Hamilton equation of matrix theory, the following relation holds good,

$$(37) \qquad e_\alpha{}^r e_\beta{}^\alpha e_s{}^\beta = \delta_s{}^r I_3 - e_s{}^r I_2 + e_\alpha{}^r e_s{}^\alpha I_1$$

and therefore

$$(38) \qquad \begin{aligned} e_\alpha{}^r e_\beta{}^\alpha e_\gamma{}^\beta e_s{}^\gamma &= \delta_\alpha{}^r e_s{}^\alpha I_3 - e_\alpha{}^r e_s{}^\alpha I_2 + e_\alpha{}^r e_\beta{}^\alpha e_s{}^\beta I_1 \\ &= \delta_s{}^r I_1 I_3 + e_s{}^r(I_3 - I_1 I_2) + e_\alpha{}^r e_s{}^\alpha (I_1{}^2 - I_2) \end{aligned}$$

and similarly in respect of higher terms. This enables us to write

$$(39) \qquad T_s{}^r = F^*{}_0(I)\delta_s{}^r + F_1(I)e_s{}^r + F_2(I)e_\alpha{}^r e_s{}^\alpha$$

and also

$$(40) \qquad T_s{}^r = F^*{}_0(I')\delta_s{}^r + F_1(I')e'_s{}^r + F_2(I')e'_\alpha{}^r e'_s{}^\alpha$$

where $F^*{}_0$, F_1 and F_2 are functions different from those of (39).

Resolving the stress, we get

$$(41) \qquad \begin{aligned} T &= F^*{}_0(I') - 2I'_2/3 \cdot F_2(I') = F_0(I') \\ T'_s{}^r &= F_1(I')e'_s{}^r + F_2(I')(e'_\alpha{}^r e'_s{}^\alpha + 2T'_2/3 \cdot \delta_s{}^r) \end{aligned}$$

where we have made use of the relation

$$(42) \qquad e'_\beta{}^\alpha e'_\alpha{}^\beta = -2I'_2.$$

In accordance with (22) and (23) we find

$$(43) \qquad \dot{w}_V = 3eF_0(I'),$$

$$(44) \qquad \dot{w}' = -2I'_2 F_1(I') + 3I'_3 F_2(I'),$$

where we have made use of the relation

$$(45) \qquad e'_\beta{}^\alpha e'_\gamma{}^\beta e'_\alpha{}^\gamma = 3I'_3.$$

If we compare (41) with (30) and (31), making use of the relation

$$e_a{}^r e_s{}^a = \zeta_s{}^r I_3 - \delta_s{}^r \mathrm{I}_2 + e_s{}^r I_1 \tag{46}$$

which is analogous to (37), we see that a dissipation function exists if F_0, F_1 and F_2 can be expressed as the derivatives of a single function as follows:

$$\begin{aligned} F_0(I') &= \partial\chi/\partial I'_1 \\ F_1(I') &= -\partial\chi/\partial I'_2 \\ F_2(I') &= \partial\chi/\partial I'_3. \end{aligned} \tag{47}$$

Elimination of χ from (47) gives

$$\begin{aligned} \partial F_0/\partial I'_2 + \partial F_1/\partial I'_1 &= 0 \\ \partial F_1/\partial I'_3 + \partial F_2/\partial I'_2 &= 0 \\ \partial F_2/\partial I'_1 - \partial F_0/\partial I'_3 &= 0. \end{aligned} \tag{48}$$

5. Expression of strain through stress. It is obvious that we can proceed quite analogously by starting from the reverse of (11), expressing $e_s{}^r$ as a function of $T_s{}^r$. We get instead of (41), (43) and (44) the following:

$$\begin{aligned} e &= \mathfrak{J}_0(T') \\ e'_s{}^r &= \mathfrak{J}_1(T')T'_s{}^r + \mathfrak{J}_2(T')\,(T'_a{}^r T'_s{}^a + 2T'_2/3\cdot\delta_s{}^r), \end{aligned} \tag{49}$$

$$\begin{aligned} \dot{w}_V &= 3T\mathfrak{J}_0(T') \\ \dot{w}' &= -2T'_2\mathfrak{J}_1(T') + 3T'_3\mathfrak{J}_2(T') \end{aligned} \tag{50}$$

where the meaning of T', T'_2 and T'_3 is obvious. If we introduce $e'_s{}^r$ from the second of (49) into the expression for $T'_s{}^r$ (second of (41)), taking into account the stress analogies to (37) and (38), we find

$$\begin{aligned} T'_s{}^r = {} & [2I'_2/3\cdot F_2 + 2T'_2/3\cdot\mathfrak{J}_2F_1 + (2T'_2/3)^2\mathfrak{J}_2{}^2F_2 \\ & + 2\mathfrak{J}_1\mathfrak{J}_2F_2(T_3 - TT_2 + 2T^3)]\delta_s{}^r \\ & + [\mathfrak{J}_1F_1 + 4T'_2/3\cdot\mathfrak{J}_1\mathfrak{J}_2F_2 + 2\mathfrak{J}_1\mathfrak{J}_2F_2(3T^2 - T_2) \\ & + \mathfrak{J}_2{}^2F_2(T_3 - TT_2 + 2T^3)]T'_s{}^r + [\mathfrak{J}_2F_1 + \mathfrak{J}_1{}^2F_2 \\ & + 4T'_2/3\cdot\mathfrak{J}_2{}^2F_2 + \mathfrak{J}_2{}^2F_2(3T^2 - T_2)]T'_a{}^rT'_s{}^a. \end{aligned} \tag{51}$$

Considering that the $\delta_s{}^r$, $T'_s{}^r$, $T'_a{}^rT'_s{}^a$ stand for the zero, first and second powers in the stress components, this requires that the first and third expressions within brackets [ ] must vanish and the second must be equal to unity. Introducing furthermore

$$\begin{aligned} T_2 &= T'_2 + 3T^2 \\ T_3 &= T'_3 + TT'_2 + T^3 \end{aligned} \tag{52}$$

we find the three equations

$$(53)\qquad \begin{aligned} &T'_2\mathfrak{F}_2F_1+[2T'^2_2/3\cdot\mathfrak{F}_2+3\mathfrak{F}_1T'_3]\mathfrak{F}_2F_2+I'_2F_2=0\\ &\mathfrak{F}_1F_1+[\mathfrak{F}_2T'_3-2T'_2/3\cdot\mathfrak{F}_1]\mathfrak{F}_2F_2=1\\ &\mathfrak{F}_2F_1+[\mathfrak{F}_1^2+T'_2/3\cdot\mathfrak{F}_2^2]F_2=0 \end{aligned}$$

from which the F_1, F_2 and I'_2F_2 can be expressed through the stress invariants. There are obviously analogous equations by which the $\mathfrak{F}_1$, $\mathfrak{F}_2$ and $T'_2\mathfrak{F}_2$ can be expressed through the strain invariants.

6. Physical interpretation. The physical interpretation of our theory will consist in assigning physical meanings to the three functions F_0, F_1, F_2 and $\mathfrak{F}_0$, $\mathfrak{F}_1$, $\mathfrak{F}_2$. For this purpose we consider certain special cases of stress and strain which constitute simple viscometric tests.

(i) The fundamental viscometric test, i e., the one which from Newton onwards served to define the coefficient of viscosity, is based upon a laminar plane deformation of the type

$$(54)\qquad e_s{}^r=e'_s{}^r=\begin{Vmatrix}0 & e_{xy} & 0\\ e_{xy} & 0 & 0\\ 0 & 0 & 0\end{Vmatrix}$$

for which

$$(55)\qquad e=0;\quad I_2=I'_2=-e_{xy}^2;\qquad I_3=I'_3=0$$

and

$$(56)\qquad e_a{}^re_s{}^a=e'_a{}^re'_s{}^a=e^2_{xy}\begin{Vmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{Vmatrix}.$$

This gives in accordance with (41)

$$(57)\quad T=F_0;\quad T'_s{}^r=F_1e_{xy}\begin{Vmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{Vmatrix}+F_2e^2_{xy}/3\begin{Vmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{Vmatrix},$$

while (43) and (44) give

$$(58)\qquad \dot{w}=\dot{w}'=2e^2_{xy}F_1.$$

Dissipation of energy, therefore, comes in this case from F_1 only. In the case of the simple Newtonian liquid the dissipated power in simple shear is $=4\eta e^2_{xy}$.[7] We follow Einstein in defining the coefficient of viscosity by equating the strain-power in our case with the strain-power of the similar flow of a simple Newtonian liquid. We accordingly get

---

[7] One should note that our e_{xy} is one half of what is usually denoted by e_{xy}.

$$(59) \qquad F_1(0, I'_2, 0) = 2\eta.$$

(ii) The simple cubical dilatation

$$(60) \qquad e_s{}^r = e\delta_s{}^r; \qquad e'_s{}^r = 0; \qquad I'_2 = I'_3 = 0$$

gives

$$(61) \qquad T_s{}^r = F_0\delta_s{}^r; \qquad \dot{w} = \dot{w}_V = 3eF_0.$$

In primitive cases rheological coefficients are defined from stress-strain relations. This method can be applied if the ratio of stress to strain is a scalar and could therefore not be applied in case (i). However, in our present case we get

$$(62) \qquad T_s{}^r/e_s{}^r = F_0(e, 0, 0)/e = \eta_V,$$

where η_V may be called the *coefficient of volume-viscosity.* The theoretical existence of such a second coefficient of viscosity was pointed out by Reiner. Recently Tisza has confirmed its actual existence from experiment.

Analogously, if we start from the simple isotropic stress $T^*{}_s{}^r = T^*\delta_s{}^r$, we arrive at a *coefficient of volume-fluidity* defined by

$$(63) \qquad \phi_V = \mathfrak{F}_0(T, 0, 0)/T.$$

(iii) The stress-analogy to (i) is the simple shearing stress in vacuo

$$(64) \qquad T^*{}_s{}^r = T^{*\prime}{}_s{}^r = \begin{Vmatrix} 0 & T^*{}_{xy} & 0 \\ T^*{}_{xy} & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}$$

so that in accordance with (8)

$$(65) \qquad \begin{aligned} &T_s{}^r = T^*{}_s{}^r - 3\kappa\epsilon\delta_s{}^r \\ &T = -3\kappa\epsilon; \quad T'_s{}^r = \end{aligned} \begin{Vmatrix} 0 & T^*{}_{xy} & 0 \\ T^*{}_{xy} & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}$$

and

$$(66) \qquad T'_2 = -T^{*2}_{xy}; \qquad T'_3 = 0$$

Expressing the strain through the stress we find from (49)

$$(67) \qquad \begin{aligned} &e = \mathfrak{F}_0 \\ &e'_s{}^r = \mathfrak{F}_1 T^*{}_{xy} \end{aligned} \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix} + \mathfrak{F}_2 T^{*2}_{xy}/3 \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{Vmatrix}.$$

We now arrange our experiment so that we keep $T^*{}_{xy}$ constant and start our observations *after a steady state has been reached.* We should then, in accordance with (67), observe, besides a continuous shearing of the liquid the rate of which is $\mathfrak{F}_1 T^*{}_{xy}$, also a continuous expansion (or contraction) the rate of which is $\mathscr{F}_0$ and a continuous lengthening (or shortening) in the direction normal to the plane of shear (in our case the z-direction) the rate of which is $\mathscr{F}_2 T^{z*}_{xy}$. The second and the third phenomenon are against experience. There may be a cubical dilatation when T^*_{xy} is applied—but this raises an elastic resistance which ultimately stops it and when the steady state is reached the rate of dilatation will have become zero. And there cannot be any extension at all in the z-direction, because, the liquid being devoid of rigidity, there would be nothing to stop it and it would go on indefinitely—which has never been observed. Therefore we must have

$$\begin{aligned} \mathscr{F}_0(0, T'_2, 0) &= 0 \\ \mathscr{F}_2(0, T'_2, 0) &= 0 \end{aligned} \tag{68}$$

which reduces (67) to

$$e_s{}^r = \mathscr{F}_1 T^*_{xy} \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}. \tag{69}$$

From (69) we arrive at the coefficient of fluidity ϕ

$$\phi = 2e_{xy}/T^*_{xy} = \mathscr{F}_1(0, T'_2, 0). \tag{70}$$

To express the stress through the strain we use (53) from which we find

$$F_2(0, I'_2, 0) = 0; \qquad F_1(0, I'_2, 0) = 1/\mathscr{F}_1(0, T'_2, 0) \tag{71}$$

so that

$$T^*{}_s{}^r = (3\kappa\epsilon + F_0)\delta_s{}^r + T^*_{xy} \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}. \tag{72}$$

This brings us back to our starting point (64) and therefore

$$3\kappa\epsilon + F_0 = 0 \tag{73}$$

or

$$\epsilon_a{}^a = 3\epsilon = -F_0(0, I'_2, 0)/\kappa. \tag{74}$$

While, therefore, in the simple Newtonian liquid a simple shearing stress produces a simple shearing strain continuously increasing at a constant rate—and nothing else; in the generalized Newtonian liquid *it will*, in addition, *produce a cubical dilatation* the magnitude of which depends upon a rheological coefficient

$$\delta = F_0(0, I'_2, 0)/I'_2 \tag{75}$$

which we may call the "coefficient of dilatancy" δ (not to be confused with Kronecker's $\delta_s{}^r$), so that

(76) $$\epsilon_a{}^a = -(\delta/\kappa)\, I'_2.$$

Conversely, *if the liquid is contained between rigid walls* as in most viscometers, so that the dilatation is prevented, *a pressure is exerted upon the walls* which is, in accordance with (57),

(77) $$T = F_0 = \delta I'_2.$$

7. Conclusions. For the purpose of our present investigations we may assume that F_2 vanishes identically and that F_0 and F_1 are independent of I'_3.[8] This results in a special case of the generalized Newtonian liquid which we may call the *Reynolds Liquid.* The rheological equation of the Reynolds liquid is accordingly

(i) $$T^*{}_s{}^r = (3\kappa\epsilon + \eta_v e + \delta I'_2)\delta_s{}^r + 2\eta e'_s{}^r$$

where $T^*{}_s{}^r$ is the total stress—elastic and viscous—3ϵ the cubical dilatation, e the isotropic component and $e'_s{}^r$ the deviator of the velocity-strain and κ, η, η_v and δ are four rheological coefficients, viz. the bulk modulus, the ordinary (Newtonian) coefficient of viscosity, the coefficient of volume viscosity and the coefficient of dilatancy, respectively. These are, to a first approximation, rheological constants with temperature and density as parameters. The total strain-power is

(ii) $$\dot{w} = (3\kappa\epsilon + \eta_v e + \delta I'_2)3e - 4\eta I'_2.$$

The contributions from free and bound energy can be separated by means of the condition that for the latter $\dot{w} \geqq 0$. We accordingly get

(iii)[9] $$\begin{aligned} \rho\dot{\phi} &= 9\kappa\epsilon e + 3\delta I'_2 e, \\ \rho\dot{\psi} &= 3\eta_v e^2 - 4\eta I'_2. \end{aligned}$$

If both coefficients of viscosity are constants, or if the coefficient of volume viscosity is a function of e only and the Newtonian coefficient of viscosity a function of I'_2 only, there exists a dissipation function

(iv) $$\chi = (3/2)\cdot e^2\eta_v - 2\eta I'_2 = \rho\dot{\psi}/2.$$

Dilatancy as defined here does not contribute to the dissipation of energy. It emerges from (i) in two forms, viz.:

(i) A deviator of velocity strain for which ϵ and e vanish produces an isotropic pressure $= \delta I'_2$.

---

[8] I'_3 and F_2 appear in what Trouton called "viscous traction." This will be dealt with in a separate communication.

[9] I'_2 is always negative.

(ii) A deviator of stress for which T^* vanishes produces a cubical dilation which ultimately (i. e., when $e = 0$) $= -(\delta/\kappa) \cdot I'_2$.

It is the latter phenomenon which Reynolds described as "dilatancy."

I take occasion to thank Professor Racah for the interest which he took in the preparation of the paper.

THE STANDARDS INSTITUTION OF PALESTINE,
TEL AVIV.

BIBLIOGRAPHY

Birch, F., *Journal of Applied Physics*, vol. 9 (1938), p. 279.
Eddington, A. S., *The mathematical theory of relativity*, Cambridge, 1924, p. 53.
Einstein, A., *Ann. d. Physik* (4), vol. 19 (1906), p. 289.
Hamel, G., *Elementare Mechanik*, Berlin, 1922.
Hencky, H., *Ann. d. Physik* (5), vol. 2 (1929), p. 617.
Hersey, M. D., *Rheology*, vol. 3 (1932), p. 196.
Maxwell, J. C., *Phil. Trans.*, vol. 157 (1867), p. 49.
Murnaghan, F. D., *American Journal of Mathematics*, vol. 59 (1937), p. 235.
Rayleigh, Lord (J. W. Strutt), *Proceedings of the London Mathematical Society*, vol. 4 (1873), p. 357.
Reiner, M., *Ten Lectures on Theoretical Rheology*, Jerusalem, 1943.
Reynolds, O., *Phil. Mag.* (5), vol. 20 (1885), p. 469.
Tisza, L., *Phys. Rev.*, vol. 61 (1942), p. 531.
Trouton, F., *Proceedings of the Royal Society* (A), vol. 77 (1906), p. 426.
Weissenberg, K., "Die Mechanik deformierbarer Körper," *Abh. Akad. d. Wiss.*, Berlin, 1931.

ALGEBRAISATION OF PLANE ABSOLUTE GEOMETRY.*

By Y. Why Tschen.

Introduction. We give in this paper a simple method, based on the axioms given by K. Reidemeister and F. Bachmann,[1] of deriving the algebraisation of plane absolute geometry up to the existence of a quadratic form. On studying the paper AG just mentioned and the well-known paper of J. Hjelmslev, "*Neue Begründung der ebenen Geometrie,*"[2] we thought it might be possible on the one hand to generalize Hjelmslev's halfrotations to those with an improper point as center and on the other hand to find out some metric properties of the generalized halfrotations. Indeed we have only to formulate Hjelmslev's halfrotations in a slightly different manner to obtain the generalization. A very useful lemma for us concerning the metric properties of the halfrotations may be stated as follows: "*If a halfrotation $H_{\mathfrak{d}}$ which belongs to the rotation $\mathfrak{d}$*[3] *transforms a line l into a line l', then the halfrotation $H_{\mathfrak{d}^{-1}}$ which belongs to the inverse-rotation $\mathfrak{d}^{-1}$ will transform any line orthogonal to l' into a line orthogonal to l.*" This lemma can easily be proved.

To prove Desargues' and Pascal's theorems we shall not use Hessenberg's method of "involutions of equal angles"[4] by which Hjelmslev proved Pascal's theorem containing only proper elements. Nor do we need the artificial concept of "cinematic space" (AG § 4) of Reidemeister and Bachmann. We shall first prove a special case of Desargues' theorem: the Desargues' configuration with any proper point P as perspective center and its polar p as perspective axis. The same theorem about the configurations with any proper point $Q \neq O$ as center but always with the polar o of the same proper point O as axis will be reduced to the special case just mentioned by using our lemma. Finally the most general Desargues' configuration which contains the line o and eventually improper points and lines will be transformed by half-

---

* Received November 22, 1944.

[1] F. Bachmann, "Eine Begründung der absoluten Geometrie in der Ebene," *Mathematische Annalen*, vol. 113, pp. 424-451. This paper will be referred to as AG.

F. Bachmann and K. Reidemeister, "Die metrische Form in der absoluten und der elliptischen Geometrie," *Mathematische Annalen*, vol. 113, pp. 748-765.

[2] *Mathematische Annalen*, vol. 64, pp. 449-474. We shall refer to this paper as NG.

[3] $\mathfrak{d}$ must be non-involutoric. The center of $\mathfrak{d}$ may be proper or improper.

[4] G. Hessenberg, *Grundlagen der Geometrie* (1930), p. 79.

rotations with the center O[5] into those configurations which, save for o, contain only proper elements. Thus all Desargues' configurations containing the fixed line o exist. Hence, according to a well-known theorem, Desargues' theorem is generally true in the projective plane.

We prove the special case of Desargues' theorem mentioned above by showing the existence of perspective collineations[6] with P as center and p as axis. We obtain these collineations as the product transformations of halfrotations and their inverses with the center P.[7] From the fact that halfrotations with the same proper point as center are commutative[8] we incidentally obtain Pascal's theorem (which is equivalent to the commutativity of perspective collineations with the same point as center).

We prove the existence of a polarity in our geometry with the help of our lemma. The involution of the right angles at a proper point can be induced by the product transformation of a halfrotation and an inverse halfrotation. As the latter are linear transformations the involution is projective and, consequently, the polarity is a linear correlation. Thus we obtain a quadratic form which is left invariant by the motions in the plane.

Our whole development is based on the axioms given in AG § 1 and their immediate consequences in AG § 2. We shall also use the results in AG, § 3 concerning improper points. We consider, before we start, that the improper points have been introduced into the plane. So if we say "points" we mean proper or improper points. But we shall have nothing to do with the improper lines, before we have introduced them. Although we follow the same principle as in AG and NG to introduce the improper lines, our method of representing the totality of points on an improper line is quite different; and our proof of the existence of a polarity depends partially upon the possibility of such representations.

We sketch in the last section a foundation of hyperbolic geometry by adding to the axioms of plane absolute geometry a new axiom, which guarantees the existence of boundary points. This new axiom does not involve any notions not previously used. Properties of boundary points are studied and improper lines introduced in a more explicit way.

---

[5] NG, p. 471.

[6] See for reference Veblen and Young, *Projective Geometry* I (1916), p. 72. The existence of such collineations with two homologous arbitrarily given points is equivalent to the existence of all Desargues' configurations with center P and axis p.

[7] We come to this product transformation by considering the Cayley formula $S = (1 + R)/(1 - R)$ which connects an orthogonal matrix R with a skew matrix S. A half-rotation can be represented as $\frac{1}{2}(1 + R)$ and the product of two skew matrices is a diagonal matrix, which represents a perspective collineation.

[8] AG, Theorem 63; NG, p. 466.

The question as to what extent and under what conditions the semi-euclidean and semi-noneuclidean geometries exist is not studied in this paper.

Notations. Capital Latin letters denote points, small Latin and Greek letters denote lines as well as reflexions. A point, which is common to two different lines a and b will also be denoted by (a, b) and a line, which is determined by two different points A, B of which at least one is proper will be denoted by $< A, B >$. If $a \perp b$, then (a, b) is a proper point.

I. Halfrotations.

Definition 1. A motion in the plane, which can be represented by $a \cdot \alpha \cdot \alpha'$ with $\alpha \perp a$, $\alpha' \perp a$ is called a sliding-reflexion. A sliding-reflexion with $\alpha = \alpha'$ is simply a reflexion.

THEOREM 1. *Let a motion $\mathfrak{M}$ be a sliding-reflexion represented by $a \cdot \alpha \cdot \alpha'$. Then the mid-point of a proper point P and its image $\mathfrak{M}(P)$ always lies on a.*

Proof. If P is on a, the theorem is trivial. Let P be not on a, and let $\bar{\alpha}$ be the line passing through P and orthogonal to a. Then we have $\alpha \cdot \alpha' = \bar{\alpha} \cdot \bar{\alpha}'$ with $\bar{\alpha} \perp a$, $\bar{\alpha}' \perp a$. By the motion $\mathfrak{M}$ the point P will be transformed into its image $\mathfrak{M}(P) = a \cdot \bar{\alpha} \cdot \bar{\alpha}'(P) = \bar{\alpha} \cdot a \cdot \bar{\alpha}'(P) = a \cdot \bar{\alpha}'(P)$ for P lies on $\bar{\alpha}$ and $\bar{\alpha}(P) = P$. We see immediately that $(a, \bar{\alpha}')$ is the mid-point of P and $\mathfrak{M}(P)$.

THEOREM 2. *If a sliding-reflexion can be represented both by $a \cdot \alpha \cdot \alpha'$ and by $b \cdot \beta \cdot \beta'$ with $\alpha \perp a$, $\alpha' \perp a$; $\beta \perp b$, $\beta' \perp b$ then the two lines a and b are not distinct and $\alpha \cdot \alpha' = \beta \cdot \beta'$.*

Proof. We have, in the previous theorem, given the line a a geometrical meaning as the locus of midpoints. Hence $a = b$ and $\alpha \cdot \alpha' = \beta \cdot \beta'$.

Definition 2. The line a is called the axis of the sliding-reflexion $a \cdot \alpha \cdot \alpha'$.

THEOREM 3. *The product of any three reflexions is a sliding-reflexion.*

Proof. If the three lines pass through a point then their product is a reflexion. Otherwise we may choose three lines a', b' and c' so that $a' \cdot b' \cdot c' = a \cdot b \cdot c$ and (a', b') is a proper point. Let $\bar{b}$ be the perpendicular to c' which passes through (a', b') and $\bar{a} \cdot \bar{b} = a' \cdot b'$. Then we have $\bar{a} \cdot \bar{b} \cdot c' = \bar{a} \cdot (\bar{b} \cdot c') = \rho \cdot (d \cdot \bar{\rho})$ where we denote by d the line which passes through $(\bar{b}, c')$ and is orthogonal to $\bar{a}$; $\rho \equiv \bar{a}$, and ρ' is the line orthogonal to d at $(\bar{b}, c')$. Hence $a \cdot b \cdot c$ is the sliding-reflexion $d \cdot \rho \cdot \rho'$.

Definition 3. Halfrotations. Let $\mathfrak{d}$ be a rotation in the plane with $\mathfrak{d}^2 \neq 1$. The center of $\mathfrak{d}$ may be proper or improper. For each line x in the plane we form the product $x \cdot \mathfrak{d}$ which can be represented as a sliding-reflexion $y \cdot \eta \cdot \eta'$, with $\eta \perp y$, $\eta' \perp y$. The axis y is uniquely determined by x and $\mathfrak{d}$. The line transformation $x \rightarrow y$ is called the halfrotation $H_{\mathfrak{d}}$. In case $\mathfrak{d} = a \cdot b$ we use also the symbol $H_{a.b}$ instead of $H_{\mathfrak{d}}$.

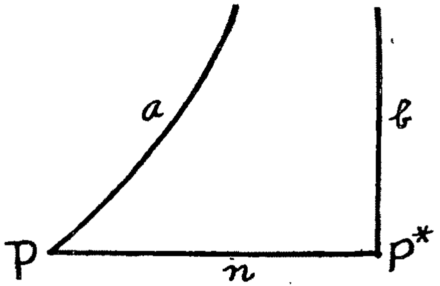
THEOREM 4. *If $x \neq x'$, then $H_{\mathfrak{d}}(x) \neq H_{\mathfrak{d}}(x')$.*

Proof. In case $x \cdot \mathfrak{d} = y \cdot \eta \cdot \eta'$ and $x' \cdot \mathfrak{d} = y \cdot \bar{\eta} \cdot \bar{\eta}'$ with $x \neq x'$ we would have $x \cdot x' = (x \cdot \mathfrak{d})(\mathfrak{d}^{-1} \cdot x') = (y \cdot \eta \cdot \eta')(\bar{\eta}' \cdot \bar{\eta} \cdot y) = \eta \cdot \eta' \cdot \bar{\eta}' \cdot \bar{\eta}$ with the η's all orthogonal to y. Hence $x \perp y$, $x' \perp y$ and $x \cdot \mathfrak{d} = y \cdot \eta \cdot \eta'$ would imply $\mathfrak{d} = x \cdot y \cdot \eta \cdot \eta' = y \cdot \eta''$ with $\eta'' \perp y$, which means $\mathfrak{d}^2 = 1$. As we have always $\mathfrak{d}^2 \neq 1$ in $H_{\mathfrak{d}}$, so $H_{\mathfrak{d}}(x) \neq H_{\mathfrak{d}}(x')$.

THEOREM 5. *Let P be any proper point and let P^* be the mid-point of P and $\mathfrak{d}(P)$, $\mathfrak{d}^2 \neq 1$. Then the lines passing through P will be transformed* (1, 1) *to the lines through P^* by the halfrotation $H_{\mathfrak{d}}$.*

Proof. Let $\mathfrak{d} = a \cdot b$. We can assume that the line a passes through P. We denote by n the line that passes through P and is perpendicular to b, $n \neq a$. Then $P^* = (n, b)$ and P^* is a proper point.

Now let p be any line on P. We form $p \cdot \mathfrak{d}$ by putting $p \cdot a = \pi \cdot n$ and we have $p \cdot \mathfrak{d} = (\pi \cdot n) \cdot b = \pi \cdot (n \cdot b) = \pi \cdot p^* \cdot \pi'$ with $p^* \perp \pi$ and p^* passing through P^* and $p^* \perp \pi'$ at P^*. Hence $H_{\mathfrak{d}}(p) = p^*$.



Conversely let p^* be any line on P^*. π' denotes the line $\perp$ to p^* at P^* and π denotes the line which passes through P and is orthogonal to p^*. Then we get $p^* \cdot \pi \cdot \pi' = \pi \cdot n \cdot b$. Let $p = \pi \cdot n \cdot a$; then p is the line which is transformed by $H_{\mathfrak{d}}$ to p^*.

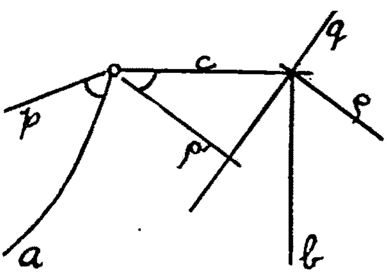
COROLLARY. *$H_{\mathfrak{d}}$ induces a point transformation among the proper points. If the pencil of lines at the proper point P is transformed into the pencil of lines at the proper point P^*, then we define $H_{\mathfrak{d}}(P) = P^*$.*

THEOREM 6. *$H_{\mathfrak{d}}$ transforms every proper point into a proper point and every line into a line. If $X \neq Y$, then $H_{\mathfrak{d}}(X) \neq H_{\mathfrak{d}}(Y)$, and if $x \neq y$ then $H_{\mathfrak{d}}(x) \neq H_{\mathfrak{d}}(y)$. The fact that the proper point X lies on the line x implies that $H_{\mathfrak{d}}(X)$ lies on $H_{\mathfrak{d}}(x)$.*

THEOREM 7. (Lemma on the metric properties of halfrotations.) *Let* $H_{\mathfrak{d}}(p) = q$. *Then the halfrotation* $H_{\mathfrak{d}^{-1}}$ *transforms every line orthogonal to* q *into a line orthogonal to* p *and if a perpendicular to* p *is the image of a line in the halfrotation* $H_{\mathfrak{d}^{-1}}$, *then the latter is a perpendicular to* q.

Proof. Suppose that p and q both pass through the center of $\mathfrak{d}$; then $\mathfrak{d} = p \cdot q$, $\mathfrak{d}^{-1} = q \cdot p$. According to the definition of $H_{\mathfrak{d}^{-1}}$ any line $\rho \perp q$ will be transformed into the line which passes through (q, ρ) and is orthogonal to p.

We now consider the case where p and q do not pass through the center of $\mathfrak{d}$. Let ρ be any perpendicular to q and let $(q, \rho) = Q$. We can assume that $\mathfrak{d} = a \cdot b$ with b on Q, and $b \neq q$. Let c be the line $\perp b$ at Q; then $c \neq a$, as $\mathfrak{d}^2 \neq 1$. The assumption $H_{\mathfrak{d}}(p) = q$ implies the existence of a perpendicular ρ' to q so that $p \cdot \mathfrak{d} = p \cdot a \cdot b = q \cdot \rho' \cdot \rho$. Since $q \cdot \rho = b \cdot c$, we have $p \cdot a \cdot b = q \cdot \rho' \cdot \rho = \rho' \cdot c \cdot b$. Hence $p \cdot a = \rho' \cdot c$ or $p \cdot \rho' = a \cdot c$. Therefore the lines p and ρ' both pass through the proper or improper point (a, c). $\rho \cdot \mathfrak{d}^{-1} = \rho \cdot b \cdot a = q \cdot c \cdot a = q \cdot \rho' \cdot p$. We denote (q, ρ') by Q' and the line, which passes through Q' and is orthogonal to p, by π; then we have $q \cdot \rho' \cdot p = \pi \cdot p' \cdot p$ with $p' \perp \pi$ at Q'. Hence $\rho \cdot \mathfrak{d}^{-1} = \pi \cdot p' \cdot p$ with $p \perp \pi$, $H_{\mathfrak{d}^{-1}}(\rho) = \pi$.



Finally let $\pi \perp p$ and suppose that there exists a line x with $H_{\mathfrak{d}^{-1}}(x) = \pi$. We take a proper point X on x and let $H_{\mathfrak{d}^{-1}}(X) = Y$. Then Y lies on π. The perpendicular to q which passes through X must be transformed by $H_{\mathfrak{d}^{-1}}$ into the perpendicular to p, which passes through Y. The latter is π. Hence $x \perp q$.

Definition 4. A halfrotation $H_{\mathfrak{d}}$ is called ordinary, if the center of $\mathfrak{d}$ is a proper point. This point will be called the center of $H_{\mathfrak{d}}$.

THEOREM 8. *The line transformation in the plane, defined by an ordinary halfrotation with the center* O, *can be stated as follows:*[9]

1° *If* g *passes through* O, $g \cdot g_1 = \mathfrak{d}$, *then* $H_{\mathfrak{d}}(g) = g_1$.

2° *If* g *does not pass through* O, *then let* $q \perp g$ *where* q *passes through* O, $q \cdot q_1 = \mathfrak{d}$. *The line* g_1 *which passes through* (q, g) *and is orthogonal to* q_1 *is the image of* q *in* $H_{\mathfrak{d}}$.

---

[9] This was used in AG as the definition of halfrotations. The proof of this is immediate.

From this it follows immediately, that any line orthogonal to a line a on O will be transformed to a line $\perp b$, if $H_{\mathfrak{h}}(a) = b$.

THEOREM 9. *An ordinary halfrotation* $H_{\mathfrak{h}}$ *transforms a proper point into a proper point, but the image of an improper point in* $H_{\mathfrak{h}}$ *may be improper or proper. The point transformation* $H_{\mathfrak{h}}$ *of all the points in the plane is a* $(1,1)$ *transformation. Hence an ordinary* $H_{\mathfrak{h}}$ *always possesses an inverse point transformation* $H_{\mathfrak{h}}^{-1}$. *If* $H_{\mathfrak{h}}(x) = x'$ *then the totality of points on* x *will be transformed into the totality of points on* x'. *In case there exists for a given line* y *a line* y', *which has the line* y *as its image in* $H_{\mathfrak{h}}$, *we shall write* $H_{\mathfrak{h}}^{-1}(y) = y'$. *Then the totality of points on* y *will be transformed by* $H_{\mathfrak{h}}^{-1}$ *into the totality of points on* y'.

Proof. The fact that all the lines which pass through the same improper point X are transformed by an ordinary $H_{\mathfrak{h}}$ into the lines all of which pass through a proper or an improper point Y is proved in AG, Theorem 58. From this it follows that an ordinary halfrotation defines a transformation of proper and improper points $H_{\mathfrak{h}}(X) = Y$.[10] We shall prove that this point transformation is $(1,1)$ in the plane. With regard to the incidence relations between points (proper and improper) and lines an ordinary halfrotation is a collineation. Two distinct points P and Q must have distinct images. For proof we take a proper point $R \neq P$, $R \neq Q$, not on $< P, Q >$, and the lines $< R, P > = p$, $< R, Q > = q$. Let $P = (p, p')$, $Q = (q, q')$. Then the lines p' and q' do not pass through R. $H_{\mathfrak{h}}(p) \neq H_{\mathfrak{h}}(q)$. $H_{\mathfrak{h}}(p')$ does not pass through $H_{\mathfrak{h}}(R)$. In case $H_{\mathfrak{h}}(P) = H_{\mathfrak{h}}(Q)$, all the images $H_{\mathfrak{h}}(p)$, $H_{\mathfrak{h}}(q)$, $H_{\mathfrak{h}}(q')$ and $H_{\mathfrak{h}}(p')$ must pass through the same point $H_{\mathfrak{h}}(R)$, which is impossible.

Now let P be any point of the plane. As $H_{\mathfrak{h}}$ always transforms proper points into proper points, so we can arbitrarily choose two proper points P_1 and P_2 in the plane which are images of proper points P^*_1 and P^*_2 respectively, and which do not lie on the same line passing through P. We may assume $P = (p_1, p_2)$ with p_1 on P_1 and p_2 on P_2. According to Theorem 5 there exist lines p^*_1 and p^*_2 with $H_{\mathfrak{h}}(p^*_1) = p_1$, $H_{\mathfrak{h}}(p^*_2) = p_2$. Hence the point $P^* = (p^*_1, p^*_2)$ is transformed by $H_{\mathfrak{h}}$ into P.

Thus the point transformation $H_{\mathfrak{h}}$ is proved to be $(1,1)$.

Definition 5. An improper point is called the pole of a line l, if that point is determined by the pencil of perpendiculars to l.

THEOREM 10. *Let* P *be any improper point, which is not the pole of a*

---

[10] See Corollary to Theorem 5.

line which passes through the proper point O. Then there exists an ordinary $H_{\mathfrak{h}}$ with the center O which transforms P into a proper point.

Proof. Let $< O, P > = p$, $P = (p, p')$. We denote the line $\perp p'$ and passing through O by n. Then $H_{p.n}(P) = (p' n)$.

THEOREM 11. *Let $H_{\mathfrak{h}}$ and $H_{\mathfrak{h}'}$ be two ordinary halfrotations with the same center O. Then $H_{\mathfrak{h}'}^{-1} \cdot H_{\mathfrak{h}}(P) = H_{\mathfrak{h}} \cdot H_{\mathfrak{h}'}^{-1}(P)$ for any point P.*

Proof. See AG, Theorem 63 and NG, p. 466.

As an application to the theorems concerning ordinary halfrotations, we prove the following theorem of which we shall make use later on.

THEOREM 12. *Let a, b, c be three lines all passing through the same proper point O and let the lines a and b be not orthogonal to c. Then there exists a line l which cuts a and b in two distinct proper points and $l \perp c$.*

Proofs. Let B be any proper point on b, $B \neq O$. We denote the line through B and $\perp c$ by n and we denote (n, c) by C. Whether $(a, n) = A$ is proper or improper, A is certainly not the pole of any line passing through O. Hence there exists an ordinary halfrotation $H_{\mathfrak{h}}$ which transforms the point A into a proper point and the line n into a line l'. By applying the halfrotation $H_{\mathfrak{h}}$ to l' we get the line l as required.

If we denote $H_{\mathfrak{h}}(A)$ by A', $H_{\mathfrak{h}}(B)$ by B', and $H_{\mathfrak{h}}(C)$ by C', then we have the relation $H_{a.c}(A') = C' = H_{b.c}(B')$. This fact will also be used later in Theorem 38.

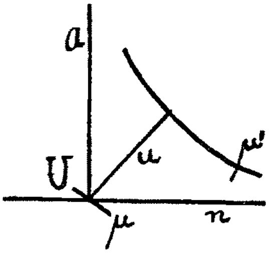
II. Improper Lines; the O-lines.

Definition 6. The totality of the points in the plane, which satisfy the following condition will be denoted by $< a; U >$ where U denotes a point lying on a. A point X belongs to $< a; U >$ if, and only if, for any two distinct lines x and y passing through X, the rotation $x \cdot y$ can be represented as the product transformation of the reflexion a and a sliding-reflexion whose axis u passes through U. In symbols the condition may be stated as follows: for $x \neq y$, $X = (x, y)$, $x \cdot y = a \cdot u \cdot \mu \cdot \mu'$ with $\mu \perp u$, $\mu' \perp u$ and u passing through U.

The point U is contained in $< a; U >$ by putting $\mu = \mu'$, and the pole of a is also contained in $< a; U >$ as is easily verified by setting in the formula $u = a$.

THEOREM 13. *Let U be a proper point on a. Then $< a; U >$ is the line orthogonal to a at the point U.*[11]

Proof. Let n be the line $\perp a$ at U. If X is a point on n and $X = (x, y)$, then $x \cdot y = n \cdot \mu'$, where μ' denotes a line on X. u is the line which passes through U and is orthogonal to μ' and $\mu \perp u$ at U. Then we have $n \cdot \mu' = n \cdot a \cdot a \cdot \mu' = a \cdot (n \cdot a)\mu' = a \cdot u \cdot \mu \cdot \mu'$. Hence X belongs to $< a; U >$.



Conversely let $x \cdot y = a \cdot u \cdot \mu \cdot \mu'$ with u passing through U and $\mu \perp u$, $\mu' \perp u$. We can assume that μ passes through U. Then $a \cdot (u \cdot \mu)\mu' = a \cdot (n \cdot a) \cdot \mu' = n \cdot \mu'$. Hence (x, y) lies on n.

THEOREM 14. *If U is an improper point, then all the points on $< a; U >$ are improper.*

Proof. We prove that if a proper point P is contained in $< a; U >$, then $< a; U >$ is the line passing through P orthogonal to a at U. Hence U is proper.

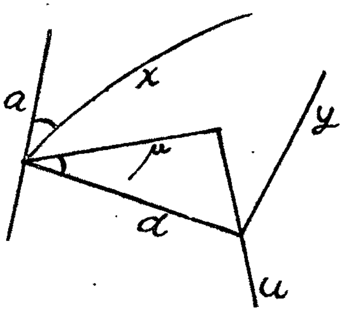
Let p be the line which passes through P and is orthogonal to a. Then we may put $P = (p, y)$. Since P lies on $< a; U >$, we have $p \cdot y = a \cdot u \cdot \mu \cdot \mu'$, with u on U and $\mu \perp u$, $\mu' \perp u$. From $a \cdot p \cdot y = u \cdot \mu \cdot \mu'$ we see at once that u passes through (a, p). Hence $U = (a, p)$.

THEOREM 15. *Let there be a configuration of six lines x, y, a, d, u and μ with the properties: 1°. The four lines x, a, d and μ are concurrent at a proper point O with $x \cdot d = a \cdot \mu$. 2°. The line $y \perp d$ and (d, y) is distinct from O. 3°. The line u passes through (d, y) and is orthogonal to μ. Any halfrotation or inverse-halfrotation with O as center will transform such a configuration into a configuration with the same properties; by this we mean that, on denoting the images of the six lines by x', y' b, d', z and λ, respectively, we have: 1°. The four lines x', b, d' and λ are concurrent at O with $x' \cdot d' = b \cdot \lambda$. 2°: $y' \perp d'$ and $(d', y') \neq O$. 3°. $z \perp \lambda$ and λ passes through (d', y'). For the inverse-halfrotation it must be assumed that the image y' of y really exists.*

Proof. This follows immediately from Theorems 8 and 9. That the images of x, a, μ and d in an inverse halfrotation really exist, follows from the fact that a halfrotation with the proper point O as center induces a $(1, 1)$ line-transformation in the pencil of lines at O. If y' exists, then (d', y') and, hence, z exist. Finally it may be mentioned that in the configuration the line x cannot be orthogonal to a, for otherwise v would coincide with d.

---

[11] By this we mean that $< a; U >$ is the totality of points lying on the line $\perp a$ at U.



COROLLARY. *Let* μ' *denote the line* $\perp u$ *at* (d, y) *and let* λ' *be the line* $\perp z$ *at* (d', y'). *The two configurations of six lines may be interpreted as* $x \cdot y = a \cdot u \cdot \mu \cdot \mu'$ *and* $x' \cdot y' = b \cdot z \cdot \lambda \cdot \lambda'$. *Our theorem states, that* $x \cdot y = a \cdot u \cdot \mu \cdot \mu'$ *implies* $x' \cdot y' = b \cdot z \cdot \lambda \cdot \lambda'$ *where* x', y', b, z *and* λ *are the images of* x, y, a, u *and* μ, *respectively, in a halfrotation or an inverse halfrotation.*[12]

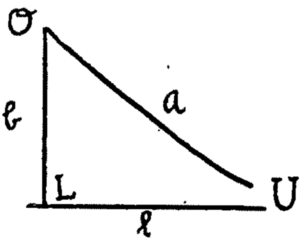
THEOREM 16. *The ordinary halfrotation* $H_{a.b}$ *transforms* $< a; U >$ *into* $< b; Z >$ *if* $H_{a.b}(U) = Z$. *The inverse halfrotation* $H_{a.b}^{-1}$ *with the proper point* (a, b) *as center transforms* $< b; U >$ *into* $< a; Z >$ *if* $H_{a.b}^{-1}(U) = Z$.

This follows immediately from the corollary to the above theorem and from the remark that, with $H_{a.b}(a) = b$, the pole of a is transformed into the pole of b by $H_{a.b}$.

THEOREM 17. *Let* U *be a proper or an improper point determined by two different lines* a *and* l *and let* O *be any proper point, distinct from* U, *on* a. *Let* b *be the line passing through* O *and orthogonal to* l. *Then the halfrotation* $H_{a.b}$ *transforms* $< a; U >$ *into the line* l, *if* U *is not the pole of any line on* O.

Proof. The assumption that U is not the pole of b, means that a and b are not orthogonal. Hence $H_{a.b}$ exists. Since $H_{a.b}(U) = (b, l) \equiv L$, we have $< a; U > \rightarrow < b; L >$ by $H_{a.b}$. By Theorem 13 $< b; L >$ is exactly the line l.

Remark. In case U is proper, this theorem follows from Theorem 8.



THEOREM 18. *Consider an ordinary halfrotation* $H_{\mathfrak{h}}$ *with the center* O.

---

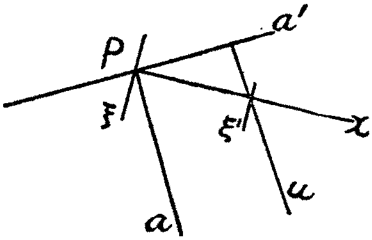
[12] The line λ' is not necessarily the image of μ'. This is true only when the plane is singular-absolute. See V. Theorem 52.

Any totality of points in the plane which can be transformed by $\Pi_{\mathfrak{d}}$ *into a given line* l *can be represented by* $< a; U >$, *where* a *is a line passing through* O *and* U *is not the pole of any line on* O.

Proof. Let b be the line passing through O and $\perp l$, $a = \mathfrak{d} \cdot b$ and $U = (a, l)$. Then $H_{a \cdot b}(U) = (b, l)$. Since $\mathfrak{d}^2 \neq 1$ a is not orthogonal to b and U is not the pole of b.

Definition 7. The polar p of a proper point P is defined as the totality of the poles of the lines which pass through P.

THEOREM 19. *Let* a *and* a' *be any two mutually orthogonal lines at the proper point* P *and let* U *be the pole of* a'. *Then the polar* p *of* P *is* $< a; U >$.



Proof. p contains the pole U of a' as well as the pole of a since p is the polar of P.

Let now x be any line which passes through P, $x \neq a$, $x \neq a'$ and $\xi \perp x$ at P, $\xi' \perp x$, $\xi' \neq \xi$. Let u be the line which passes through (x, ξ') and is $\perp$ to a'. Then $\xi \cdot \xi' = (\xi \cdot x) \cdot (x \cdot \xi') = a \cdot a' \cdot u \cdot \mu$ with $\mu \perp u$ at (x, ξ'). Since $a' \perp u$, we have $(\xi, \xi') =$ the pole of x which is contained in $< a; U >$.

Conversely, let (ξ, ξ') be any point contained in $< a; U >$. We may assume that ξ passes through P. Then $\xi \cdot \xi' = a \cdot u \cdot \mu \cdot \mu'$ with $u \perp a'$. Hence $\xi \cdot \xi' = a \cdot a' \cdot u \cdot \mu'' \perp u$. From this it follows that ξ and ξ' are the perpendiculars of the line connecting P and (μ'', u). Hence (ξ, ξ') lies on p.

THEOREM 20. $< a; U >$ *is the polar of a proper point if, and only if,* U *is the pole of a line.*

Proof. If U is the pole of a line, this line must be orthogonal to a. Our theorem follows immediately from Theorem 19.

We are now ready to extend our plane by introducing the so-called "improper O-lines," where O is a proper point arbitrarily chosen in the plane.

Definition 8. What we have up to the present called "lines" will be in future called "proper lines." Let O be a proper point in the plane, then an improper O-line is any $< a; U >$, where a is a proper line which passes through O and U is an improper point on a. Proper and improper O-lines will be also called O-lines. The proper and improper points together with the O-lines form the O-plane.

Lying on an improper O-line there is no proper point. Hence if we know that a line l passes through a proper point, then this line l is certainly proper. We may then occasionally omit the adjective "proper."

If a and b are two distinct lines passing through O, and U (on a) and Z (on b) are both contained in the polar o of O, then $< a; U >$ as well as $< b; Z >$ represent the polar o.

Definition 9. Two O-lines are said to be distinct if the totality of the points lying on the one is not the same as the totality of the points lying on the other.

An improper O-line is always distinct from a proper line, for on an improper line there is no proper point.

THEOREM 21. *Two improper O-lines $< a; U >$ and $< b; Z >$ are not distinct if, and only if,* 1) $a = b$, $U = Z$, 2) *In $< a; U >$ and $< b; Z >$ the points U and Z are contained in o.*

Proof. If one of the two points U and Z, say U, is not contained in o, then $< a; U >$ is distinct from o, and $< a; U > \neq b, Z' = o$. Hence the two O-lines are distinct. If both U and Z are not contained in o, then we have the cases 1, $a = b$; 2, $a \neq b$. In case $a = b$, $U \neq Z$ the two O-lines $< a; U >$ and $< a; Z >$ are distinct because the only point common to a and $< a; U >$ is U, as is immediately seen from the formula $x \cdot y = a \cdot u \cdot \mu \cdot \mu'$. In case $a \neq b$ and $< a; U > = < b; Z >$, the improper O-line $< a; U >$ must contain the pole of a as well as the pole of b. Hence with $\beta \perp b$, $\beta' \perp b$, $\beta' \neq \beta$ we have $\beta \cdot \beta' = a \cdot u \cdot \mu \cdot \mu'$ with $u \neq a$ and u lying on U. But the pole of b is also contained in o, and if $a' \perp a$ at O and W is the pole of a', then $o = < a; W >$ and $W \neq U$. We have then also $\beta \cdot \beta' = a \cdot w \cdot \lambda \cdot \lambda'$ with w on W. Hence $u \cdot \mu \cdot \mu' = w \cdot \lambda \cdot \lambda'$. It follows from Theorem 2 that $u = w$ and $U = W$ which is impossible.

THEOREM 22. *Aside from the polar o of O each O-line has an unique representation as $< a; U >$ where a denotes a line which passes through O and U is a point on a but not contained in o.*

THEOREM 23. *Any improper O-line distinct from o can be transformed by a suitably chosen halfrotation with the center O into a proper line.*

THEOREM 24. *A halfrotation H or an inverse halfrotation H^{-1} with the center O induces a* $(1, 1)$ *transformation in the O-lines. Together with the H or H^{-1} as a* $(1, 1)$ *point transformation in the plane, we see that H or H^{-1} is a* $(1, 1)$ *collineation in the O-plane.*

Proof. H or H^{-1} transforms every point on o into a point on o; hence H or H^{-1} leaves the polar o fixed. Every point U not on o has in H or H^{-1} an unique image Z not on o. The line-transformation induced in the pencil of lines at O is $(1,1)$. From $H(a) = b$ or $H^{-1}(a) = b$ (with a, b on O) and $H(U) = Z$ or $H^{-1}(U) = Z$ we have immediately $< a; U > \rightarrow < b; Z >$ by H or H^{-1}.

THEOREM 25. *In the O-plane the points and the O-lines satisfy the incidence-relations:* 1) *two distinct O-lines have one and only one point in common;* 2) *there is one and only one O-line which passes through two distinct points.*

Proof. With a product of halfrotations with O as center we can always transform two points not on o or two O-lines $\neq o$, respectively, into two proper points and two proper lines. Our theorem follows from the fact, that the halfrotations and their inverses are collineations in the O-plane. The cases where one of the given points is on o or the polar o is one of the given O-lines, require a special discussion which can be furnished without any difficulty.

Remark. For every proper point P there is an extended P-plane obtained by introducing the improper P-lines. We do not know so far whether for any two distinct proper points O and P the extended P-plane is the same as the extended O-plane. We shall prove this fact as in AG by applying Desargues' theorem.

THEOREM 26. *Let P and O be two distinct proper points. Then the polar p of P is always an improper O-line.*

Proof. We have $O \neq P$; let $a = < O; P >$, n be the line $\perp a$ at P. Then we have $p = < a; N >$, where N denotes the pole of n.

III. Theorems of Desargues and Pascal.

THEOREM 27. *Let P be any proper point in the plane whose polar is p. All the Desargues' configurations with the point P as perspective center and the polar p as perspective axis are true in the extended P-plane.*

Proof. We shall prove, in the extended P-plane, the existence of all the perspective collineations with P as center, p as axis.

Let a be any proper line on P. A and A' are two points arbitrarily given on a with the condition that $A \neq P$ and $A' \neq P$ do not both lie on p. We shall construct the perspective collineation which leaves P and p fixed and transforms A into A'. As is well known, such a collineation, if it exists, will be uniquely determined by A and A'.

Let $b \perp a$ at P and let B be a proper point on b, $P \neq B$. Then the lines $d = < B, A >$ and $d' = < B, A' >$ are both proper and not orthogonal to b nor to a. Let $n \perp d$ and n through P, $n' \perp d'$ and n' also through P. Then n and n' are both distinct from a and b and are not orthogonal to them. Hence we can form the ordinary halfrotations $H_{a.n} \cdot H_{b.n}$ and their inverses $H_{a,n}^{-1}$ and $H_{b.n'}^{-1}$. The product transformation $K = H_{a.m} \cdot H_{b,m}^{-1} \cdot H_{b,n'} \cdot H_{a.n'}^{-1}$ is a collineation in the P-plane, which leaves P and p fixed. Since $K(a) = a$, K transforms every line on P into itself. But $H_{a.n} \cdot H_{b.n}^{-1}(A) = B$, $H_{b.n'} \cdot H_{a.n'}^{-1}(B) = A'$; hence $K(A) = A'$. Then K is the required collineation.

From this theorem follows a special form of Desargues' configuration, which besides P, p and the points on p consists only of proper points and proper lines. These special cases may be stated as follows:

THEOREM 28. *Let P be any proper point in the plane, and let r, s, t be three different proper lines passing through P. Let R_1 and R_2, S_1 and S_2, P_1 and T_2 be proper points lying on r, s and t, respectively, and all distinct from P. If now*

$$< R_1, S_1 > \perp \tau, \qquad < R_2, S_2 > \perp \tau;$$
$$< S_1, T_1 > \perp \rho, \qquad < S_2, T_2 > \perp \rho;$$
$$< T_1, R_1 > \perp \sigma,$$

where ρ, τ, σ are three proper lines passing through P, then $< T_2, R_2 > \perp \sigma$.

We notice that in the above statement no mention is made of the improper P-lines. Therefore this theorem can be regarded as a theorem in the original plane.

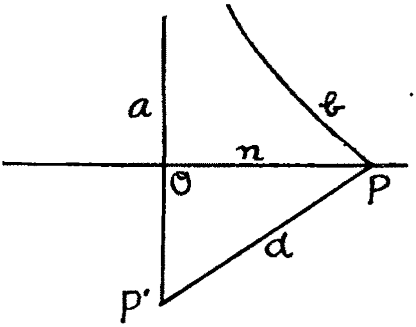
We now take a proper point O in the plane and extend the latter to the O-plane. We proceed to prove the validity of the general Desargues' theorem in the O-plane, i. e., Desargues' configurations which consist of points and O-lines.

THEOREM 29. *Let P' be any proper point and r', s' and t' be any three distinct lines passing through P'. Let R'_1 and R'_2, S'_1 and S'_2, T'_1 and T'_2 be proper points lying on r', s' and t', respectively, and all distinct from P'. If now*

$$< R'_1, S'_1 > \perp \tau', \qquad < R'_2, S'_2 > \perp \tau';$$
$$< S'_1, T'_1 > \perp \rho', \qquad < S'_2, T'_2 > \perp \rho';$$
$$< T'_1, R'_1 > \perp \sigma',$$

where σ', ρ', τ' *are three proper lines which pass through the proper point* O, *then* $< T'_1, R'_2 > \perp \sigma'$.

Proof. If $P' = O$, then this theorem follows from the previous one by putting $P \equiv O$.



We may assume $P' \neq O$. We denote $< OP' >$ by a and the line $\perp a$ at O by n. We take a proper point $P \neq O$ on n and denote d by $< P, P' >$ and the line $\perp d$ at P by b. Certainly $a \neq b$ and $(a \cdot b)^2 \neq 1$. We form the halfrotation $H_{a.b}$ which may be ordinary or not.

We shall prove that $H_{a.b}$ transforms the configuration stated in this theorem into a configuration which is described in Theorem 28.

According to Theorem 5 the halfrotation $H_{a.b}$ transforms $P' \rightarrow P$, $r' \rightarrow r$, $s' \rightarrow s$, $t' \rightarrow t$ and $R'_1 \rightarrow R_1$, $R'_2 \rightarrow R_2$, $S'_1 \rightarrow S_1$, $S'_2 \rightarrow S_2$, $T'_1 \rightarrow T_1$, $T'_2 \rightarrow T_2$ where r, s and t are three lines passing through P and R_1 and R_2, S_1 and S_2, T_1 and T_2 are six proper points distinct from P and lying on r, s, and t, respectively.

We now consider instead of $H_{a.b}$ the halfrotation $H_{b.a}$. We have $H_{b.a}(P) = O$. Hence, again according to Theorem 5, $H_{b.a}$ transforms the lines at P (1, 1) into the lines at O. Therefore there exist three lines σ, ρ and τ which pass through P so that $H_{b.a}(\rho) = \rho'$, $H_{b.a}(\sigma) = \sigma'$ and $H_{b.a}(\tau) = \tau'$.

By our lemma (Theorem 7) we see, incidentally, that the perpendiculars of σ', ρ' and τ' will be transformed into the perpendiculars of σ, ρ and τ respectively. Hence we have:

$$\begin{array}{ll} < R_1, S_1 > \perp \tau, & < R_2, S_2 > \perp \tau; \\ < S_1, T_1 > \perp \rho, & < S_2, T_2 > \perp \rho; \\ < T_1, R_1 > \perp \sigma. & \end{array}$$

According to Theorem 28 $< T_2, R_2 > \perp \sigma$ must hold. As $< T_2, R_2 >$ is the image of $< T'_2, R'_2 >$ in $H_{a.b}$, it follows from the lemma that $< T'_2, R'_2 > \perp \sigma'$.

The above theorem can be stated as a theorem in the O-plane as follows. Any Desargues' configuration in the O-plane, which has a proper point as perspective center and the polar of O as perspective axis and which consists, besides the three improper points on o, of proper points only, is true.

THEOREM 30. *All the Desargues' configurations in the* O*-plane which contain the polar* o *of* O, *are true.*

Proof.[13] First let the perspective center of the configuration be outside o. As such a configuration contains only a finite number of improper points which do not lie on o, we can transform them into proper points by a product of halfrotations. The original configuration will be transformed into a configuration described in Theorem 29. Hence the latter configuration is true. By the inverse process of halfrotations we see, incidentally, the truth of the general configuration.

If the center of the configuration lies on o its validity is shown by an indirect proof.

THEOREM 31. *Desargues' theorem is true in the O-plane.*

THEOREM 32. *Pascal's theorem is true in the O-plane.*

Proof. We prove first a special case of Pascal's theorem. As constructed in Theorem 27 every perspective collineation with center O and axis o is a product of halfrotations and inverse-halfrotations with O as center. Since the halfrotations with the same proper point as center are commutative it follows at once that the perspective collineations with the center O are commutative. From this follows immediately a special case of Pascal's theorem.

The general case will be proved by using the translations in the O-plane.

Having established the validity of Desargues' theorem in the O-plane we can use it to prove that, for any two different proper points O and O', the totality of the points on an improper O-line is the same as the totality of the points on an improper O'-line. Hence the extension of the plane obtained by introducing improper lines does not depend upon the particular proper point O chosen as center.

A proper O-line is also a proper O'-line. The polar of a proper point is an improper O-line as well as an improper O'-line. Given any improper O-line s, which is not a polar, we shall prove the existence of Desargues' configurations which contain besides s only proper lines and polars. As the Desargues theorem holds in the O-plane as well as in the O'-plane, it follows readily that s is also an improper O'-line.

Given an improper O-line s, which is not a polar of any proper point, let o denote the polar of O and let $(o \cdot s) = S$. Let A_1 and A_2 be any two different points on s, each distinct from S. We write $a_1 = < O, A_1 >$, $a_2 = < O, A_2 >$. According to Theorem 12 there exists a proper line s' which cuts a_1 and a_2 in two proper points A'_1 and A'_2 and which passes through S. Let a_3 be another proper line on O, distinct from a_1 and a_2, and let A_3 be a proper point on a_3, $A_3 \neq O$; let $< A_3, A_1 > = r$, $< A_3, A_3 > = t$; then r and t are proper lines. We connect A'_1 and the point (r, o) with the line r', A'_2 and (t, o) with t'. Then we have a Desargues configuration which consists of 8 proper lines a_1, a_2, a_3, t, r, s', r', t' and the polar o and the improper O-line s. From this it follows that A_1, A_2 and S also lie on the same improper O'-line. If we vary the point A_2 and keep A_1 and S fixed, we see at once that s is an improper O'-line.

---

[13] The proof follows NG, p. 471.

THEOREM 33. *The extension of the plane obtained by introducing improper lines does not depend upon the particular proper point O chosen as center.*

We shall, in the following, use the terms "improper lines" instead of "improper O-lines" and "lines" instead of "O-lines," the "extended plane" instead of "O-plane."

THEOREM 34. *Let P be any proper point in the plane. Any line in the extended plane can be represented as $< a; U >$ with a passing through P.*

Proof. We can regard the extended plane as the P-plane.

THEOREM 35. *Any motion in the plane induces a collineation in the extended plane.*

Proof. A motion $\mathfrak{M}$ is always a $(1, 1)$ point transformation in the extended plane, because $a \cdot b \cdot c = d$ implies $\mathfrak{M}(a) \cdot \mathfrak{M}(b) \cdot \mathfrak{M}(c) = \mathfrak{M}(d)$. If, now, an improper line is represented by $< a; U >$ and $\mathfrak{M}(a) = a'$, $\mathfrak{M}(U) = U'$, then $\mathfrak{M}$ transforms $< a; U >$ into $< a'; U' >$, since relations of the form $\chi \cdot y = a \cdot u \cdot \mu \cdot \mu'$ must also hold among the image lines.

IV. Polarity. Quadratic Form.

Definition 10. If there exists a rectangle in the plane, the geometry is called singular absolute, otherwise ordinary absolute.[14]

THEOREM 36. *All the proper points in a singular absolute geometry have the same polar—the absolute polar.*

Proof. By assumption there exists a rectangle formed, say, by the four proper lines γ, γ', δ and δ' with the $\gamma, \gamma' \perp \delta, \delta'$. Then if γ'' is a third proper line $\gamma'' \perp \delta$, the product $\gamma \cdot \gamma' \cdot \gamma''$ is a reflexion (AG, Theorem 20). Hence, according to Theorem 21 in AG, γ'' is also orthogonal to δ' as χ and γ' are.

---

[14] According to the definition given by Reidemeister and Bachmann.

Thus the two proper lines γ and γ' have the same pencil of perpendiculars, which we denote by $\mathfrak{P}_1$. Since the same holds true for δ and δ', we may denote the pencil of their common perpendiculars by $\mathfrak{P}_2$. Thus we have an orthogonal net in the plane. Through every proper point there are two mutually orthogonal lines, the one belonging to $\mathfrak{P}_1$, the other to $\mathfrak{P}_2$. All the proper lines in $\mathfrak{P}_1$ have the same pole P_1 and all the proper lines in $\mathfrak{P}_2$ have the same pole P_2.

Thus the polar of any proper point passes through P_1 and P_2. Since $P_1 \neq P_2$, this polar is uniquely determined by P_1 and P_2. Hence all the proper points have the same polar.[15]

THEOREM 37. *In an ordinary absolute geometry different proper points have always different polars.*

THEOREM 38. *To every improper point U, which is not the pole of a proper line, there exists uniquely an improper line ω—called the polar of the improper point U*[16] *—with the property, that any proper line, which passes through U, has its pole lying on ω.*

Proof. We take a proper line a which passes through U, and on this line take a proper point O. We can form the improper line $< a; U >$ which is not the polar of any proper point (Theorem 22).

Let $u_1 \neq a$ be any proper line passing through U and $\beta_1 \perp u_1$, β_1 on O. Since U is not the pole of a proper line $(a \cdot \beta_1)^2 \neq 1$ and $H_{\beta_1 . a}$ exists. This halfrotation $H_{\beta_1 . a}$ will transform the pole of u_1 into a point, say Z_1. Let β'_1 be another perpendicular of u_1, $\beta'_1 \neq \beta_1$, and let $H_{\beta_1 . a}(\beta'_1) = z_1$; then we have evidently $Z_1 = (z_1, a)$ since $H_{\beta_1 . a}(\beta_1) = a$. According to our lemma the halfrotation $H_{a . \beta_1}$ will transform the pole of z_1 into the pole of β'_1. But the pole of β'_1 is a point on u_1 and so the pole of z_1 is a point belonging to $< a; U >$.

If we take another proper line u_2 which passes through U and $u_2 \neq a$, $u_2 \neq u_1$ and let β_2 be the perpendicular of u_2 through O, then the halfrotation $H_{\beta_2 . a}$ will transform the pole of u_2 into a point, say, Z_2. And if β'_2 is another perpendicular of u_2, $\beta'_2 \neq \beta_2$ and $H_{\beta_2 . a}(\beta'_2) = z_2$ then evidently we have $Z_2 = (z_2, a)$. And as proved above for z_1, the proper line z_2 has its pole lying on $< a; U >$.

According to Theorem 12 there exist two proper points P_1 and P_2 with P_1 not on β_1 and P_2 not on β_2, so that $H_{\beta_1 . a}(P_1) = P = H_{\beta_2 . a}(P)$. We may assume that β'_1 passes through P_1 and that β'_2 through P_2. Then z_1 as well as z_2 must pass through the same proper point P. If $z_1 \neq z_2$, then $< a; U >$, which contains the poles of z_1 and z_2, would be the polar of the proper point P. This is impossible. Hence $z_1 = z_2$ and $(z_1, a) = Z_1 = Z_2 = (z_2, a)$.

---

[15] Another simple proof of this theorem, which does not make use of the extended plane, is given in the Appendix.

[16] Up to the present we have only defined the polar of a proper point (Definition 7).

It follows that whatever the proper line u_i may be, provided u_i passes through U, the halfrotation $H_{\beta_i . a}(\beta_i \perp u_i \beta_i \text{ on } O)$ always transforms the pole of u_i into the same point, which we denote by Z. Or in symbols we have

$$\beta'_i \cdot (\beta_i \cdot a) = z_i \lambda\lambda' \quad \text{with} \quad \lambda \perp z_i,\ \lambda' \perp z_i,\ z_i \quad \text{on} \quad Z.$$
$$\beta_i \cdot \beta'_i = a \cdot z_i \cdot \lambda' \cdot \lambda.$$

In words: the pole (β_i, β'_i) of u_i lies on $< a; Z >$. If we put $\omega = < a; Z >$ then ω is the line required.

If there were two distinct lines ω, ω' with the same property, then all the proper lines, which pass through U would have the same point (ω, ω') as pole. Then U would be the pole of any proper line orthogonal to the u_i. This contradicts our assumption.

THEOREM 39. *To every improper line ω, which is not the polar of a proper point, there exists uniquely an improper point Z—called the pole of the improper line ω—with the property, that the proper lines, which pass through Z, and only these, have their poles lying on ω.*

Proof. We denote ω by $< a; U >$, where U is an improper point, which is not the pole of any proper line. Starting from U we construct, as in Theorem 38, the point $Z = (z_i, a)$, so that the pole of z_i lies on $< a; U > = \omega$. Since the pole of a is certainly contained in $< a; U >$, we have at least two different proper lines a and z_i which pass through Z and have their poles lying on ω.

Z must be improper for otherwise ω would be the polar of a proper point. Hence either Z is the pole of a proper line or Z is not the pole of any proper line. In the latter case we apply Theorem 38 to the improper point Z, from which it follows, that all the poles of the proper lines passing through Z must lie on the same line. Hence this line must be ω. Therefore Z can be defined as the pole of ω.

In case Z is the pole of a proper line l, then the poles of a and z_i must be the same point (ω, l). Hence all the proper lines on Z have this point as their pole.

Any proper line m which does not pass through Z, can not have its pole lying on ω, for m must intersect some proper lines on Z in proper points.

By applying the above two theorems in the singular absolute and ordinary absolute geometries we shall prove the existence of an involution and a polarity.

THEOREM 40. *Let the geometry be singular absolute. The polar of an improper point, which is not the pole of a proper line, is the absolute polar* π. *The pole of an improper line* ω, *which is not the absolute polar, is the same as the pole of proper lines which pass through* (ω, π). *Hence follows the existence of an involution, a* $(1, 1)$ *point transformation on* π *induced by the right angles at any proper point. If in this involution* A *and* A' *are a pair of corresponding points, then* $A \neq A'$ *and the lines passing through* A *have* A' *as pole and those passing through* A' *have* A *as pole.*

Proof. For the proof it is sufficient to remark, that any two different proper lines p and p' may have the same pole or different poles according as the point (p, p') is itself a pole or not. In the latter case the polar of (p, p') must be the absolute polar.

THEOREM 41. *Let the geometry be ordinary absolute. To every point (proper or improper!) there exists a line called its polar and to every line (proper or improper!) there exists a point called its pole. Different points have different polars, different lines have different poles. If* P *has* p *as polar, then* p *has* P *as pole.*

Proof. For the proof it is sufficient to remark, that different proper lines have always different poles.

THEOREM 42. *The polarity in an ordinary absolute geometry is a correlation and a reciprocity.*

This theorem is equivalent to

THEOREM 43. *If a point* P *lies on a line* l, *then the polar* p *of* P *passes through* L, *the pole of* l.

In order to prove this theorem, we need the following generalization of our lemma:

THEOREM 44. *Let* $H_{\mathfrak{h}}$ *be an ordinary halfrotation. If* $H_{\mathfrak{h}}(P) = P'$, *then* $H_{\mathfrak{h}^{-1}}$ *transforms the polar* π' *of* P' *into the polar* π *of* P. *Conversely if* $H_{\mathfrak{h}}$ *transforms a line* π *into a line* π', *then* $H_{\mathfrak{h}^{-1}}$ *transforms the pole* P' *of* π' *into the pole* P *of* π.

Proof. From $H_{\mathfrak{h}}(P) = P'$, it follows that every proper line p passing through P will be transformed into a proper line p', which passes through P'. Hence, according to our lemma, $H_{\mathfrak{h}^{-1}}$ transforms the pole of p' into the pole of p. Since $H_{\mathfrak{h}^{-1}}$ is a collineation on the extended plane, and $H_{\mathfrak{h}^{-1}}$ transforms more than one point on π' into points on π, so $H_{\mathfrak{h}^{-1}}(\pi') = \pi$.

Conversely if $H_{\mathfrak{h}}(\pi) = \pi'$, then $H_{\mathfrak{h}^{-1}}(P') = P$. For otherwise from $H_{\mathfrak{h}^{-1}}(P') = P''$ it would follow that $H_{\mathfrak{h}}(\pi'') = \pi'$. Hence $\pi'' = \pi$ and $P'' = P$.

Proof of Theorem 43. Let P be a point on l, and let l be a proper line. The polar π of P is defined as the line containing all the poles of proper lines passing through P. Since l is such a proper line, we see at once that the pole L of l must be contained in π; in other words π passes through L.

Let now l be an improper line. Then there always exists an ordinary halfrotation $H_{\mathfrak{h}}$ which transforms l into a proper line l'. $H_{\mathfrak{h}}(l) = l'$, $H_{\mathfrak{h}}(P) = P'$. P' lies on l'. According to the part of the theorem just proved, the polar π' of P' must pass through the pole L' of l', since l' is proper. From Theorem 44 we have $H_{\mathfrak{h}^{-1}}(\pi') = \pi$ and $H_{\mathfrak{h}^{-1}}(L') = L$. Hence π passes through L, because $H_{\mathfrak{h}^{-1}}$ is a collineation in the extended plane.

THEOREM 45. *Any motion in the plane leaves the involution on the absolute polar or the polarity in the ordinary absolute geometry invariant.*

Proof. Any motion is a collineation in the extended plane. Furthermore any motion leaves the orthogonality-relations invariant. This proves the theorem.

We now proceed to show that the polarity is not only a correlation, it is also a *projective correlation.* From this follows immediately the existence of a quadratic form.

A collineation or a correlation K is certainly projective, if it transforms a certain one-dimensional form F_1 into the same or another one-dimensional form F_2, in such a way that the mapping induced by K in the elements of F_1 and F_2 is projective, that is to say, if the mapping leaves the cross-ratios unchanged. We shall use this remark to prove the existence of a quadratic form as follows:

1°. Any reflexion in the plane is a projective transformation in the extended plane.

Proof. A reflexion leaves every point on its axis fixed; it is furthermore a collineation in the plane. Hence it is a projective transformation of period 2. From this it follows, also, that the coördinate field of our geometry is not of characteristic 2.

2°. Any rotation in the plane is a projective transformation.

3°. Any ordinary halfrotation is a projective transformation.

Proof. Suppose that the ordinary halfrotation $H_{\mathfrak{h}}$ has the proper point O as center. The polar o of O will be transformed into itself by $H_{\mathfrak{h}}$ and $\mathfrak{h}$. We introduce homogeneous coördinates $[x_1 : x_2 : x_3]$ in the plane, with the line o as $x_3 = 0$. If a proper point $P[\xi : \eta : 1]$ is transformed into the proper point $P'[\xi' : \eta' : 1]$ by $\mathfrak{h}$ and into the proper point $\bar{P}[\bar{\xi} : \bar{\eta} : 1]$ by $H_{\mathfrak{h}}$ then $\bar{P}$ is the mid-point of P and P' and $< O, \bar{P} > \perp < P, P' >$. Hence the reflexion which has the proper line $< \bar{P}, O >$ as axis transforms P into P' and the points $\bar{P}$ and the point of intersection P_0 of $< P, P' >$ with o are fixed. Therefore $PP' ; \bar{P}P_0$ form a harmonic point set. As P_0 lies on o, it must have the coördinate $[\xi - \xi' : \eta - \eta' : 0]$. Hence it follows that the coördinate of $\bar{P}$ is $[\frac{1}{2}(\xi + \xi') : \frac{1}{2}(\eta + \eta') : 1]$.

We represent the rotation $\mathfrak{h}$ in the affine coördinates (x_1, x_2) with o as the line of infinity) by a linear transformation

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = (A) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{1}$$

where (A) denotes a square matrix. From this equation we get the equations of transformation $H_{\mathfrak{h}}$ so far as the proper points (x_1, x_2) are concerned:

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = \tfrac{1}{2}[(E) + (A)] \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{2}$$

where (E) is the unit matrix. The equation (2) represents a linear transformation $\mathfrak{T}$ in the extended affine plane. So far as the proper points are concerned, the transformation $\mathfrak{T}$ and $H_{\mathfrak{h}}$ are the same. Hence $\mathfrak{T} \cdot H_{\mathfrak{h}}^{-1}$ leaves every proper point fixed. But if every proper point is a fixed point, then all the points in the extended plane are unchanged by the transformation $\mathfrak{T} H_{\mathfrak{h}}^{-1}$. Hence $H_{\mathfrak{h}} = \mathfrak{T}$, and $H_{\mathfrak{h}}$ is represented as a linear transformation (2).

THEOREM 46. *The involution of right angles at a proper point is a projective transformation.*

Proof. The involution of right angles at a proper point P is induced by the product of halfrotations $H_{a.m} \cdot H_{b.m}^{-1}$ as described in Theorem 27. Since halfrotations are linear transformations the involution is projective.

THEOREM 47. *In the singular absolute geometry there exists a quadratic form $x_1^2 + kx_2^2$ which will be left invariant by those motions in the plane which are products of reflexions. $(-k)$ is not equal to a square in the coördinate field.*

THEOREM 48. *The polarity in the ordinary absolute geometry is a linear reciprocity. Hence there exists a quadratic form $x_1^2 + k_2x_2^2 + k_3x_3^2$ with*

$k_2 \cdot k_3 \neq 0$ *which is left invariant by those motions*[17] *in the plane, which are products of reflexions.*

Definition 11. A singular absolute geometry is called euclidean, if there exist no improper points other than those on the absolute polar. When such points exist we call the geometry semi-euclidean.

Definition 12. An ordinary absolute geometry is said to be of elliptic type, if there is no point which lies on its polar. The geometry is called elliptic, if every improper point is the pole of a proper line; otherwise it is termed semi-elliptic.

Definition 13. An ordinary absolute geometry is said to be of hyperbolic type, if there exists at least one point which lies on its polar. Such points are called boundary points. The geometry is called hyperbolic, if, save for boundary points, every improper point is the pole of a proper line; otherwise the geometry is called semi-hyperbolic.

The elliptic geometry has been studied by E. Podehl and K. Reidemeister in: "Eine Begründung der elliptischen Geometrie," *Hamburger Abhandlungen*, vol. 10 (1934). It is easy to see that our method of algebraization can be applied there equally well. There is no need to introduce improper lines and things are simpler.

V. Hyperbolic Geometry.

Hyperbolic geometry will be obtained by directly introducing to the axioms given in AG, § 1 some new axioms, which specify the nature of improper points. We assume that improper points have been introduced into the absolute geometry according to AG, § 3. We shall have nothing to do with the improper lines before we define them, which can be done much more explicitly in this special case. The results developed above concerning the polarity will not be used.

We first state some theorems which can be proved easily by using the axioms of absolute geometry.

THEOREM 49. *A rotation with a proper point O as center leaves no point different from O fixed.*

THEOREM 50. *A sliding-reflexion with the line l as axis will transform a point P, which does not lie on l, into itself if and only if P is the pole of l.*

---

[17] We do not know whether every motion is a linear collineation, and in the non-euclidean case, whether every linear collineation, which leaves the quadratic form invariant, is a motion.

Proof. Let the sliding-reflexion be $l \cdot \alpha \cdot \alpha'$. If $\alpha = \alpha'$, the theorem is true. Let $\alpha \neq \alpha'$, $(\alpha, l) = A$, $< A, P > = p$, $\beta \perp p$ at A. Then $l \cdot \alpha \cdot \alpha' = p \cdot \beta \cdot \alpha'$. From the assumption $l \cdot \alpha \cdot \alpha'(P) = P$ follows $p \cdot \beta \cdot \alpha'(P) = P$ so that $\beta(P) = \alpha'(P)$; hence $\alpha' \perp p$, which is impossible unless P is the pole of l.

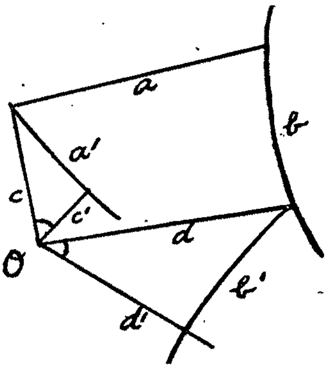
THEOREM 51. *Let α and α' be two different perpendiculars of a line l. The rotation $\alpha \cdot \alpha'$ transforms a point P not lying on l into itself if and only if P is the pole of l.*

Proof. From $\alpha \cdot \alpha'(P) = P$ follows $p \cdot \alpha \cdot \alpha'(P) = P$ if p connects (α, l) with P. But $p \cdot \alpha \cdot \alpha'$ is a sliding-reflexion with the axis passing through (α', l). By applying Theorem 50 we obtain the theorem immediately.

Axiom A. O. There is no rectangle in the plane.

THEOREM 52. *An ordinary halfrotation $H_{\mathfrak{h}}$ with O as center, will not transform two mutual orthogonal lines a, b which both do not pass through O, into mutually orthogonal lines.*

Proof. Let c be the line through O and $\perp a$, and let d be the line through O and $\perp b$. $H_{\mathfrak{h}}(c) = c'$, $H_{\mathfrak{h}}(d) = d'$, $H_{\mathfrak{h}}(a) = a'$, $H_{\mathfrak{h}}(b) = b'$. Then a' passes through (c, a) and b' passes through (b, d).



According to our lemma $H_{\mathfrak{h}^{-1}}$ will transform any line $\perp b'$ into a line $\perp b$. If $a' \perp b'$, then $H_{\mathfrak{h}^{-1}}(a') = a'' \perp b$. But $\mathfrak{h}^{-1} = c' \cdot c$, and a'' passes through (c', a') and $\perp c$. Hence a, b, c and a'' form a rectangle, which contradicts our axiom A. O.

Axiom B_1. There are on a line at most two improper points, which are not the poles of (proper) lines.

Definition 14. Improper points, which are not the poles of lines in the plane, are called, if they exist, boundary points.

THEOREM 53. *If on a line l there is a boundary point R there exist exactly two boundary points on l.*

Proof. Let $\alpha \perp l$. Then $\alpha(R) \neq R$ since R is not the pole of α. But if R is not a pole, $\alpha(R)$ is not a pole either. Hence $\alpha(R)$ is a boundary point. According to Axiom B_1, R and $\alpha(R)$ are the two boundary points on l. The improper points on l distinct from R and $\alpha(R)$ are poles.

Axiom B_2. There exists at least one boundary point in the plane.

Remark. If we take all the lines which pass through a boundary point R, then we get on each line besides R one other boundary point. All these boundary points are distinct. We shall show later, that these are exactly all the boundary points which exist in the plane.

THEOREM 54. *If R is a boundary point lying on the line r, $\alpha \perp r$, $\alpha' \perp r$, then $\alpha \cdot \alpha'(R) = R$.*

THEOREM 55. *If a and r are two lines, which pass through a boundary point R, then a rotation, which can be represented by $a \cdot r \cdot \alpha \cdot \alpha$ with $\alpha \perp r$, $\alpha' \perp r$ will leave the point R fixed.*

THEOREM 56. *Let $x \cdot y = a \cdot r \cdot \alpha \cdot \alpha'$ with a, r passing through a boundary point R; $\alpha, \alpha' \perp r$, $x \perp r'$ where r' is also a line on R, then $y \perp r'$.*

Proof. From a proper point Y on y we take the perpendicular y' of r'. Then $x \cdot y'(R) = R$ (Theorem 54). Hence $y' \cdot y = (y' \cdot x)(a \cdot r \cdot \alpha \cdot \alpha')$ leaves R fixed, and we have $y = y'$ (Theorem 49).

THEOREM 57. *Given a boundary point R and a line a passing through R. Any rotation in the plane which leaves R fixed can be represented as $a \cdot r \cdot \alpha \cdot \alpha'$ with r passing through R and $\alpha \perp r$, $\alpha' \perp r$.*

Proof. Let $\mathfrak{d}(R) = R$, then $a \cdot \mathfrak{d}(R) = R$. But $a \cdot \mathfrak{d}$ is a sliding-reflexion, hence its axis must pass through R (Theorem 50).

THEOREM 58. *Any point, which is contained in $< a; R >$ is either the boundary point itself or the pole of a line passing through R.*

Proof. Let the point P in $< a; R >$ be (x, y) with $x \cdot y = a \cdot r \cdot \alpha \cdot \alpha'$, $\alpha \perp r$, $\alpha' \perp r$, r on R. We can assume $\alpha \neq \alpha'$. To prove that there exists a line r' passing through R with $x \perp r'$, $y \perp r'$, amounts to showing that P is the pole of r'.

1°. P can not be a proper point, for $x \cdot y$ would then not leave R fixed.

2°. P can not be the pole of a line not passing through R according to Theorem 51.

3°. If P is a boundary point distinct from R, then we get a contradiction by the following argument: We take a proper point O on a and let α be $\perp r$ and pass through O. Let d denote the line connecting O and (α', r). We get x and y by the conditions: $x \cdot d = a \cdot \alpha$, $y \perp d$ at (α', r). The ordinary halfrotation $H_{x.a}$ transforms P into R and hence $< x; P >$ into $< a; R >$. Let d' be another perpendicular of y, $d' \neq d$, then (d, d') belongs to $< x; P >$ and is transformed into a point of $< a; R >$. But $H_{x.a}(d) = \alpha$. Hence according to Theorem 56 $H_{x.a}(d')$ is a line $\perp r$. This is impossible, by Theorem 52, since $H_{x.a}(y) = r$, $\alpha \perp r$, $d \perp y$, $d' \perp y$.

4°. The only possibility remaining for P is that P is the pole of a line passing through R.

THEOREM 59. *All the rotations whose centers are either the boundary point R or the pole of any line passing through R form a group: the group of all rotations in the plane which leave R fixed.*

In proving Theorem 58, 3°, we have used the fact that an ordinary halfrotation can never transform a boundary point into another boundary point. Now it is also easy to prove that an ordinary halfrotation never transforms a boundary point into a pole. If $H_{\mathfrak{h}}$ transforms a boundary point R into the pole of a line l, then $H_{\mathfrak{h}^{-1}}$ will transform l into a line l', which must have R as the pole, according to our lemma. This is impossible. Hence by an ordinary halfrotation a boundary point is always transformed into a proper point. From this follows

THEOREM 60. *Given a boundary point R and a line s not passing through R then there exists a perpendicular of s which passes through R.*

Proof. We take a proper point O on s and let $r = < O, R >$. We assume r not $\perp$ to s. $H_{r.s}$ transforms R into a proper point P. The perpendicular to s at P certainly passes through R.

THEOREM 61. *Any two distinct boundary points are connected by a line.*

Proof. Let R_1 and R_2 be two distinct boundary points, $R_1 = (a, b)$. Then we can write $R_2 = (p, q)$ with $p \perp a$, $q \perp b$ (Theorem 60). Let l be the line connecting (a, p) and (b, q), $t \perp l$ and t passing through R_1. Then the line r with $r = p \cdot t \cdot q$ connects R_2 and R_1 as is easily seen from the relation

$$a \cdot r \cdot b = (a \cdot p) \cdot t \cdot (q \cdot b) = (lt_1)t(t_2l) = t'$$

with t_1, t_2, t' all $\perp l$.

Definition 15. Improper lines in the hyperbolic geometry. There are two kinds of improper lines, 1) Polar of a proper point, 2) Tangent at a boundary point R.—This consists of R and all the poles of the lines which pass through R.

THEOREM 62. *Two distinct lines (proper or improper) have one and only one point in common. Two distinct points lie on one and only one line (proper or improper).*

Proof. This theorem may be proved by discussing the different cases with regard to the kinds of lines and points which are given. By applying Theorems 61 and 62 and the relations of orthogonality among proper lines the proof follows without any difficulty.

The proofs of Desargues' and Pascal's theorems in the hyperbolic geometry follow the same principle as in III.

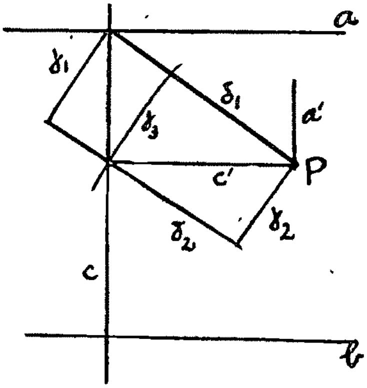
The pole-polar relation has been fixed by the definition of improper lines.

Appendix.

We give a simple proof of the theorem: *In case there exists a rectangle in the plane, then for any two different (proper) lines a and b with a common perpendicular, the pencil of lines orthogonal to a is the same as the pencil of lines orthogonal to b.*

Proof. Let the rectangle be formed by the four lines γ, γ', δ, and δ'. Then according to the proof in Theorem 36 we know, that there exists an orthogonal net in the plane consisting of two pencils of lines $\mathfrak{P}_1$ and $\mathfrak{P}_2$.

Now let a and b have a common perpendicular c. We may suppose that a and b are not contained in $\mathfrak{P}_1$ nor in $\mathfrak{P}_2$. Let γ_1 of the pencil $\mathfrak{P}_2$, and δ_1 of the pencil $\mathfrak{P}_1$ both pass through (a, c), $\gamma_1 \perp \delta_1$. We take a proper point P on δ_1, $P \neq (a, c)$. Let $c' \perp c$ and c' pass through P, and let δ_2 be the line of $\mathfrak{P}_1$ which passes through (c', c). Let γ_2 be contained in $\mathfrak{P}_2$, and on P; and let γ_3 be contained in $\mathfrak{P}_2$ and pass through (c', c). Finally, let a' be the line $\perp c'$ at P.



We have

$$c \cdot a = \gamma_1 \cdot \delta_1, \qquad a' \cdot c' = \delta_1 \cdot \gamma_2;$$

hence

$$\begin{aligned} a \cdot a' &= c \cdot \gamma_1 \cdot \gamma_2 \cdot c', \\ c \cdot a \cdot a' \cdot c &= \gamma_1 \cdot \gamma_2 \cdot c' \cdot c \\ &= \gamma_1 \cdot \gamma_2 \cdot \gamma_3 \cdot \delta_2 = \gamma_0 \cdot \delta_2 \end{aligned}$$

with $\gamma_0 \perp \delta_2$. Therefore $a \perp a'$. Thus the four lines a, a', c, c' form a rectangle. From $b \perp c$ follows $b \perp a'$.

NATIONAL SOUTH-WEST ASSOCIATED UNIVERSITY,
KUNMING, CHINA.

ON LEBESGUE SUMMABILITY AND ITS GENERALIZATION TO INTEGRALS.*

By OTTO SZÁSZ.

---

1. **Introduction.** A series $\sum_1^\infty a_n$ is called Lebesgue summable (shortly L-summable) to the value s, if the sine series

$$\sum_1^\infty n^{-1} a_n \sin nt = F(t) \tag{1.1}$$

is convergent in some interval $-\tau < t < \tau$, and if

$$t^{-1} F(t) \to s \text{ as } t \to 0; \tag{1.2}$$

we then write: $L \Sigma a_n = s$. Let

$$\gamma_n^r = \binom{n+r}{n}, \quad \text{and} \quad s_n^r = \sum_1^n \gamma_{n-\nu}^r a_\nu ; \tag{1.3}$$

the series Σa_n is summable (C, r) for some $r > -1$, to the sum s, if

$$s_n^r \sim s \gamma_n^r \text{ as } n \to \infty; \tag{1.4}$$

we then write $(C, r) \Sigma a_n = s$. Note that if r is not a negative integer, then

$$\gamma_n^r \sim \frac{n^r}{\Gamma(r+1)} \text{ as } n \to \infty. \tag{1.5}$$

Hardy and Littlewood [2, Theorem 2][1] proved that $(C, -\alpha)$ summability for some $\alpha > 0$ implies L-summability. For $\alpha = 0$ the statement no longer holds. In fact Hardy and Littlewood proved [1b, § 2], that to any positive function $\phi(n)$, tending steadily to infinity with n, it is possible to find a convergent series Σa_n, for which $a_n = O\{n^{-1}\phi(n)\}$, while

$$\limsup_{t \to 0} \sum a_n \frac{\sin nt}{nt} = \infty.$$

We replace $(C, -\alpha)$ summability by a more general assumption in the following theorem:

---

* Received October 23, 1944. Presented to the American Mathematical Society, October 28, 1944.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

THEOREM A. *If* Σa_n *is summable* $(C, 1-\alpha)$ *for some positive* $\alpha < 1$, *and if*

$$\sigma_n = \sum_1^n |s_\nu^{-\alpha}| = O(n^{1-\alpha}) \text{ as } n \to \infty, \tag{1.6}$$

then the series is L-summable.

The proof is contained in **2** and **3**. In the later sections we give several generalizations of L-summability to integrals.

2. An auxiliary theorem. The following formulae are well known:

$$(1-x)^{-r-1} = 1 + \sum_1^\infty \frac{(r+1)(r+2)\cdots(r+n)}{n!} x^n = \sum_0^\infty \gamma_n^r x^n, \tag{2.1}$$

$$\sum_0^n \gamma_\nu^r = \gamma_n^{r+1}, \qquad \gamma_n^r = \gamma_n^{r+1} - \gamma_{n-1}^{r+1}, \tag{2.2}$$

$$s_n^r = \sum_1^n \gamma_{n-\nu}^{\epsilon-1} s_\nu^{r-\epsilon}, \qquad a_n = \sum_1^n \gamma_{n-\nu}^{-r-2} s_\nu^r. \tag{2.3}$$

We first prove the following lemma:

LEMMA 1. *For* $\alpha < 2$ *the series*

$$\sum_{n=\nu}^\infty \gamma_{n-\nu}^{\alpha-2} \frac{\sin nt}{nt} = \sum_{n=0}^\infty \gamma_n^{\alpha-2} \frac{\sin(n+\nu)t}{(n+\nu)t} = T_\nu^\alpha(t) \tag{2.4}$$

is absolutely convergent, and

$$T_\nu^\alpha - T_{\nu+1}^\alpha = T_\nu^{\alpha-1}, \qquad (\nu = 1, 2, 3, \cdots). \tag{2.5}$$

Furthermore for $\alpha \leq 1$

$$|T_\nu^\alpha(t)| < 2\nu^{-1}t^{-\alpha}, \qquad t > 0, \quad \nu \geq 1. \tag{2.6}$$

Proof. If α is an integer ≤ 1, then clearly the series (2.4) is finite; if α is not an integer, then from (1.5)

$$\gamma_n^{\alpha-2} \frac{\sin(n+\nu)t}{(n+\nu)t} = O(n^{\alpha-3});$$

hence the series (2.4) is absolutely convergent for $\alpha < 2$. Furthermore, employing (2.2) we find

$$T_\nu^\alpha - T_{\nu+1}^\alpha = \frac{\sin \nu t}{\nu t} + \sum_{\nu+1}^\infty \gamma_{n-\nu}^{\alpha-3} \frac{\sin nt}{nt} = \sum_\nu^\infty \gamma_{n-\nu}^{\alpha-3} \frac{\sin nt}{nt};$$

this proves (2.5). To prove (2.6), note that from (2.1)

$$\sum_0^\infty \gamma_n^{\alpha-2} x^{n+\nu} = x^\nu (1-x)^{1-\alpha};$$

hence

$$\sum \gamma_n^{\alpha-2}\cos(n+\nu)t = \Re\, e^{i\nu t}(1-e^{it})^{1-\alpha}.$$

On putting $1-\alpha = 2\beta \geq 0$, we obtain

$$e^{i\nu t}(1-e^{it})^{2\beta} = e^{it(\nu+\beta)}(-2i\sin(t/2))^{2\beta} = e^{it(\nu+\beta)}e^{-i\beta\pi}(2\sin(t/2))^{2\beta},$$

and

$$\Re\, e^{i\nu t}(1-e^{it})^{2\beta} = (2\sin(t/2))^{2\beta}\cos\{(\nu+\beta)t-\beta\pi\}.$$

Hence

$$T_\nu^\alpha(t) = t^{-1}\int_0^t (2\sin(x/2))^{2\beta}\cos\{(\nu+\beta)x-\beta\pi\}dx;$$

if $\alpha = 1$, $\beta = 0$, $T_\nu(t) = \sin\nu t/\nu t$, and

$$|T_\nu(t)| < 1/\nu t \text{ for } t > 0, \qquad \alpha = 1, \quad \nu \geq 1.$$

If $\alpha < 1$, $\beta > 0$, an integration by parts yields

$$T_\nu^\alpha(t) = t^{-1}(2\sin(t/2))^{1-\alpha}\,\frac{\sin\{(\nu+\beta)t-\beta\pi\}}{\nu+\beta}$$
$$-(2\beta/t)\int_0^t (2\sin(x/2))^{-\alpha}\,\frac{\sin\{(\nu+\beta)x-\beta\pi\}}{\nu+\beta}\,\cos(x/2)\,dx;$$

hence

$$|T_\nu^\alpha(t)| < \frac{t^{-\alpha}}{\nu} + \frac{2\beta}{\nu+\beta}\,t^{-1}\int_0^t (2\sin(x/2))^{-\alpha}\cos(x/2)\,dx$$
$$= \frac{t^{-\alpha}}{\nu} + \frac{1}{\nu+\beta}\,t^{-1}(2\sin(t/2))^{1-\alpha} < 2\nu^{-1}t^{-\alpha}.$$

This proves (2.6), and establishes our lemma.

Remark. A. F. Andersen defines differences of fractional order of a sequence $\{\epsilon_\nu\}$ by the formula

$$\Delta^r\epsilon_\nu = \sum_{n=0}^{\infty}\gamma_n^{-r-1}\epsilon_{\nu+n};$$

hence

$$T_\nu^\alpha(t) = \Delta^{1-\alpha}\,\frac{\sin\nu t}{\nu t}.$$

We have from (2.1), for $r = \alpha - 2$,

$$\gamma_n^{\alpha-2} = (-1)^n\binom{1-\alpha}{n}.$$

3. Proof of Theorem A. Using (2.3)

$$(3.1)\quad \sum_1^{\infty} n^{-1}a_n\sin nt = \sum_{n=1}^{\infty} n^{-1}\sin nt\sum_{\nu=1}^{n}\gamma_{n-\nu}^{\alpha-2}s_\nu^{-\alpha} = \sum_{\nu=1}^{\infty}s_\nu^{-\alpha}\sum_{n=\nu}^{\infty}n^{-1}\gamma_{n-\nu}^{\alpha-2}\sin nt;$$

the interchange of summation is legitimate, as the double series is absolutely convergent for $\alpha < 1$. In fact from (1. 5)

$$n^{-1}\gamma_{n-\nu}{}^{\alpha-2}s_\nu{}^{-\alpha}\sin nt = O\{n^{-1}(n-\nu+1)^{\alpha-2}\,|\,s_\nu{}^{-\alpha}\,|\},$$

and

$$\sum_{n=\nu}^{\infty} n^{-1}(n-\nu+1)^{\alpha-2} < \nu^{-1}\sum_{1}^{\infty} k^{\alpha-2} = O(\nu^{-1})\,;$$

it remains to prove the convergence of $\sum_{1}^{\infty}\nu^{-1}\,|\,s_\nu{}^{-\alpha}\,|$. Now, employing (1. 6)

$$\sum_{1}^{n}\nu^{-1}\,|\,s_\nu{}^{-\alpha}\,| = n^{-1}\sigma_n + \sum_{1}^{n-1}\sigma_\nu(1/\nu(\nu+1)) = O(n^{-\alpha}+\sum_{1}^{n}\nu^{-1-\alpha}) = O(1)\,;$$

thus all series in (3. 1) are absolutely convergent, and

$$t^{-1}F(t) = \sum_{1}^{\infty} s_\nu{}^{-\alpha}T_\nu{}^{\alpha}(t).$$

We now choose a positive number μ, and put $n = [\mu t^{-1}]$,

$$\sum_{1}^{\infty} s_\nu{}^{-\alpha}T_\nu{}^{\alpha} = (\sum_{1}^{n}+\sum_{n+1}^{\infty})s_\nu{}^{-\alpha}T_\nu{}^{\alpha} = U_n(t)+V_n(t), \quad \text{say.}$$

From (2. 6)

$$|\,V_n(t)\,| < 2t^{-\alpha}\sum_{n+1}^{\infty}\nu^{-1}\,|\,s_\nu{}^{-\alpha}\,|,$$

and from (1. 6)

$$\sum_{n+1}^{\infty}\nu^{-1}\,|\,s_\nu{}^{-\alpha}\,| = -(\sigma_n/(n+1)) + \sum_{n+1}^{\infty}\sigma_\nu(1/\nu(\nu+1)) < \sum_{n+1}^{\infty}\sigma_\nu(1/\nu(\nu+1))$$

$$= O(\sum_{n}^{\infty}\nu^{-1-\alpha}) = O(n^{-\alpha}) = \mu^{-\alpha}t^{\alpha}O(1)\,;$$

hence

$$V_n(t) = \mu^{-\alpha}O(1).$$

Furthermore, in view of (2. 5),

$$U_n = s_n{}^{1-\alpha}T_n{}^{\alpha} + \sum_{1}^{n-1}s_\nu{}^{1-\alpha}(T_\nu{}^{\alpha}-T_{\nu+1}{}^{\alpha}) = s_n{}^{1-\alpha}T_n{}^{\alpha} + \sum_{1}^{n-1}s_\nu{}^{1-\alpha}T'_\nu{}^{\alpha-1}.$$

We may assume without loss of generality, that

$$s_n{}^{1-\alpha} = o(n^{1-\alpha})\,;$$

then, employing (2. 6)

$$U_n(t) = o(n^{-\alpha}t^{-\alpha}) + o(\sum_{1}^{n}\nu^{-\alpha}t^{1-\alpha}) = \mu^{-\alpha}o(1)+\mu^{1-\alpha}o(1).$$

Summarizing, we get

$$t^{-1}F(t) = \mu^{-\alpha}O(1) + \mu^{1-\alpha}o(1) \text{ as } t \to 0;$$

hence

$$\limsup_{t\to 0} t^{-1} \mid F(t) \mid \leq \mu^{-\alpha}O(1);$$

α being positive, μ arbitrarily large, and $O(1)$ independent of μ, we get finally

$$t^{-1}F(t) \to 0 \text{ as } t \to 0,$$

which proves our theorem. Note that the theorem becomes false for $\alpha = 0$. To illustrate the scope of the theorem, we construct a series summable $(C, 1-\alpha)$ and satisfying (1.6) but not summable (C, β) for any $\beta < 1-\alpha$. Put $s_n^{-\alpha} = k^{-1/2}2^{k(1-\alpha)}$ for $n = 2^k$, $k \geq 1$, and $s_n^{-\alpha} = 0$ otherwise. Then for $2^k \leq n < 2^{k+1}$

$$s_n^{1-\alpha} = \sum_1^n s_\nu^{-\alpha} = \sum_1^k \nu^{-1/2}2^{\nu(1-\alpha)} = o(2^{k(1-\alpha)}) = o(n^{1-\alpha});$$

hence $(C, 1-\alpha)\Sigma a_n = 0$. Moreover condition (1.6) is evidently satisfied, as $s_n^{-\alpha} \geq 0$ for all n; in fact $\sigma_n = o(n^{1-\alpha})$, which means strong summability $(C, 1-\alpha)$.

On the other hand, from (2.3) and (2.1) for $0 < \beta < \alpha < 1$

$$s_n^{\beta} = \sum_1^n \gamma_{n-\nu}^{\beta+\alpha-1} s_\nu^{-\alpha} > s_n^{-\alpha};$$

hence, for $n = 2^k$,

$$n^{-\beta}s_n^{\beta} > k^{-1/2}2^{k(1-\alpha-\beta)} \to \infty,$$

so that our series is not summable (C, β).

4. Generalization of *L*-summability to integrals. Employing Stieltjes integrals the following generalization suggests itself:

Suppose that the integral

$$\phi(t) = \int_0^\infty u^{-1} \sin ut dA(u) \text{ exists for } 0 < t < \tau, \tag{4.1}$$

and that

$$t^{-1}\phi(t) \to s \text{ as } t \to 0; \tag{4.2}$$

we then say that the integral $\int_0^\infty dA(u)$ is summable to s by the method ϕ.

If in particular $A(u)$ is a step function with the jump a_n at the point $u = n(1, 2, 3, \cdots)$, then $\phi(t) = \sum_1^\infty n^{-1}a_n \sin nt$, and our method reduces to the L-method.

If the jumps a_n occur at the points $u = \lambda_n$, where $\lambda_n \uparrow \infty$, then

$$\phi(t) = \sum_1^\infty \lambda_n^{-1} a_n \sin \lambda_n t. \tag{4.3}$$

We give now a generalization of L-summability to integrals without the use of Stieltjes integrals. By means of the identity

$$\cos(n - \tfrac{1}{2})t - \cos(n + \tfrac{1}{2})t = 2 \sin nt \sin(t/2)$$

(1.1) becomes

$$F(t) = \frac{1}{2 \sin t/2} \sum_1^\infty \frac{a_n}{n} \{\cos(n - \tfrac{1}{2})t - \cos(n + \tfrac{1}{2})t\};$$

but

$$\cos(n - \tfrac{1}{2})t - \cos(n + \tfrac{1}{2})t = t \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \sin utdu,$$

so that (1.2) is equivalent to

$$t^{-1} \sum_1^\infty n^{-1} a_n \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \sin utdu \to s. \tag{4.4}$$

We define

(4.5) $a(u) = 0$ for $0 \leq u < \frac{1}{2}$, $a(u) = un^{-1}a_n$ for $n - \frac{1}{2} \leq u < n + \frac{1}{2}$;
then (4.4) becomes

$$g(t) = t^{-1} \int_0^\infty u^{-1} a(u) \sin utdu \to s. \tag{4.6}$$

Thus the following definition also includes L-summability:

The integral $\int_0^\infty a(u)\,du$ is g-summable to s, if $\int_0^\infty u^{-1} a(u) \sin utdu$ exists in some interval $0 < t < \tau$, and if (4.6) holds.

In case of (4.5) we have the formal equation

$$\int_0^\infty a(u)\,du = \sum_1^\infty n^{-1} a_n \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} udu = \sum_1^\infty a_n.$$

5. Another generalization of *L*-summability to integrals. L-summability, as defined in **1**, is a series to function transform. We now show that it can be written as the sequence to function transform

$$\sum_1^\infty s_n \Delta \frac{\sin nt}{nt} \to s. \tag{5.1}$$

More precisely: if (1.2) holds, then the series (5.1) converges to the value $t^{-1}F(t)$, and conversely if (5.1) converges then $t^{-1}\Sigma n^{-1} a_n \sin nt$ converges to the same value.

To prove the first part, we employ a result due to B. Kuttner [3, §4], stating that (1.2) implies $s_n = o(n)$. Now

$$\sum_{1}^{n} \nu^{-1} a_\nu \sin \nu t = n^{-1} s_n \sin nt + \sum_{1}^{n-1} s_\nu \Delta \frac{\sin \nu t}{\nu}, \tag{5.2}$$

which yields the first part of our statement.

For the converse we shall use the lemma:

LEMMA 2. *If*

$$B_n(t) = a_n \sin nt + b_n \sin(n+1)t \to 0 \tag{5.3}$$

for every t belonging to a set $\mathcal{E}$ of positive measure, then $a_n \to 0$ and $b_n \to 0$.

This lemma granted, the convergence of the series (5.1) implies $s_n \Delta \dfrac{\sin nt}{nt} \to 0$; hence by the lemma $n^{-1}s_n \to 0$, and our statement follows from (5.2).

The lemma is an analogue to a well known theorem of Cantor and Lebesgue. For its proof put

$$\rho_n^2 = a_n^2 + b_n^2$$

where we may assume that $\rho_n > 0$. Clearly

$$B_n^2(t) < 2\rho_n^2;$$

hence, in view of (5.3)

$$\rho_n^{-1} B_n^2(t) < 2 \mid B_n(t) \mid \to 0 \text{ for } t \in \mathcal{E}.$$

Using a theorem of Egoroff, it follows that on a point set $\mathcal{E}' \subset \mathcal{E}$ of positive measure $\mid \mathcal{E}' \mid > 0$

$$\rho_n^{-1} B_n^2(t) \to 0 \text{ uniformly in } \mathcal{E}';$$

a fortiori

$$\rho_n^{-1} \int_{\mathcal{E}'} B_n^2(t)\, dt \to 0. \tag{5.4}$$

Now

$$2B_n^2 = a_n^2(1 - \cos 2nt) + b_n^2[1 - \cos 2(n+1)t] + 2a_n b_n[\cos t - \cos(2n+1)t];$$

hence

$$2 \int_{\mathcal{E}'} B_n^2(t)\, dt = (a_n^2 + b_n^2) \mid \mathcal{E}' \mid + 2a_n b_n \int_{\mathcal{E}'} \cos t dt + \mathcal{R}_n, \text{ say,}$$

and

$$2\rho_n^{-2} \int_{\mathcal{E}'} B_n^2(t)\, dt > \int_{\mathcal{E}'} (1 - (\cos t))\, dt - \rho_n^{-2} \mid \mathcal{R}_n \mid.$$

But $\rho_n^{-2}\mathcal{R}_n$ consists of three terms, each term being a Fourier coefficient of the characteristic function corresponding to $\mathcal{E}'$, multiplied by factors $< \pi$; hence $\rho_n^{-2}\mathcal{R}_n \to 0$ so that the sequence $\rho_n^{-2}\int_{\mathcal{E}'} B_n^2(t)\,dt$ is bounded away from zero. Combined with (5.4) this yields $\rho_n \to 0$, and our lemma is proved.

We may now write

$$t^{-1}F(t) = \sum_1^\infty s_n \Delta \frac{\sin nt}{nt} = t^{-1} \sum s_n \int_n^{n+1} u^{-2}(\sin ut - ut \cos ut)\,du;$$

let

$$A(u) = s_n \text{ for } n \leq u < n+1, \qquad (n = 1, 2, 3, \cdots),$$

then

$$F(t) = \int_1^\infty A(u) u^{-2}(\sin ut - ut \cos ut)\,du. \tag{5.5}$$

We are thus led to the following generalization of L-summability:

If the integral (5.5) exists in some interval $-\tau < t < \tau$, and if $t^{-1}F(t) \to s$ as $t \to 0$, then s is the generalized limit of $A(u)$.

The comparison of our integral transforms with Cesàro integral means will be discussed in another paper.

University of Cincinnati.

BIBLIOGRAPHY.

1. G. H. Hardy and J. E. Littlewood, "Abel's theorem and its converse," a) I, *Proceedings of the London Mathematical Society*, vol. 18 (1920), pp. 205-235; b) II, *ibid.*, vol. 22 (1923), pp. 254-269.
2. ———, "Notes on the theory of series VII: On Young's convergence criterion for Fourier series," *Proceedings of the London Mathematical Society*, vol. 28 (1928), pp. 301-311.
3. B. Kuttner, "Some relations between different kinds of Riemann summability," *Proceedings of the London Mathematical Society*, vol. 40 (1936), pp. 524-540.

ON A CONGRUENCE PROPERTY OF THE DIVISOR FUNCTION.*

By L. G. Sathe.

1. Introduction and notations. Let k be a given integer, r a fixed residue class $(\bmod k)$ and $d(n)$ the number of divisors of n. Let $N(k,r,x)$ denote the number of $n \leq x$ for which $d(n) \equiv r (\bmod k)$. We are concerned in this paper with the asymptotic behavior of $N(k,r,x)$ as $x \rightarrow \infty$. S. S. Pillai [1] has proved that $N(k,0,x) \sim Bx$ when k is a prime.[1] We here treat the problem for all values of k and r. The main result is:

Theorem. *If 2^l is the highest power of 2 that divides k, then there is a constant $B = B(k,r) > 0$ such that*

$$N(k,r,x) \sim Bx \text{ if } 2^l | r \quad \text{and} \quad N(k,r,x) = o(x) \text{ if } 2^l \nmid r$$

The lemmas are given in **2** and the main theorem is proved in **3**. In **4** we investigate under what conditions $B(k,r)$ is equal to $B(k',r')$ and evaluate the constant B in special cases.

p is a prime, $f(n)$ denotes the number of different prime factors of n, $h(n) = \prod_{p|n} \{p/(p+1)\}$, $h(1) = 1$; u denotes a number such that if $p^\theta \| u$ then either (i) $\theta = 0$ or (ii) $2 \leq \theta \leq k-1$.[2] Thus every prime divisor of u occurs at least in the second power and at most in the $(k-1)$-th power. q always denotes a square free integer.

2. Lemmas.

Lemma 1. *Let d be a given square free number and let $M(x,d,r) = M(x,d,r,e)$ be the number of square free integers $q \leq x$ for which $f(q) \equiv r \pmod e$ and $(q,d) = 1$. Then to every given positive number δ there corresponds a $\Delta = \Delta(\delta, e)$ independent of d and r such that*

$$(1) \qquad | M(x,d,r) - Ah(d)x | < \delta 2^{f(d)} x \text{ if } x \geq \Delta$$

where $A = 1/e\zeta(2)$.

---

* Received July 1, 1944. The author wishes to thank Professor F. W. Levi and Dr. S. S. Pillai.

[1] *Journal of the Indian Mathematical Society* vol. 6, N. S. (1942). The proof contains a mistake which has been pointed out and rectified by the present author, *Journal of the Indian Mathematical Society* vol. 7, N. S. (1943).

[2] $p^\theta \| n$ denotes that p^θ is the highest power of the prime p that divides n.

Proof. Let p be a given prime and let $M_1(x, r) = M_1(x, r, e)$ denote the number of $q \leq x$ for which $f(q) \equiv r \pmod{e}$. Then, obviously,

$$M(x, p, r) + M(x/p, p, r-1) = M_1(x, r). \tag{2}$$

In [2] and [3] it is proved that

$$M_1(x, r) \sim Ax \text{ for every fixed } r. \tag{3}$$

From (2) and (3) it follows that to every fixed r and arbitrarily small positive number δ there corresponds a number $\Delta' = \Delta'(\delta, e)$ independent of p and r such that

$$| M(x, p, r) - Ax + M(x/p, p, r-1) | < \tfrac{1}{2}\delta x \text{ if } x \geq \Delta'.$$

Hence,

$$\begin{aligned}
| M(x, p, r) &- Ax + Ax/p - M(x/p^2, p, r-2) | \\
&\leq | M(x, p, r) - Ax + M(x/p, p, r-1) | \\
&\quad + | M(x/p, p, r-1) - Ax/p + M(x/p^2, p, r-2) | \\
&< \tfrac{1}{2}\delta x + \tfrac{1}{2}\delta x/p \text{ if } x/p \geq \Delta'.
\end{aligned}$$

Repeating this process we get finally, if $\nu = [\log x/2 \log p]$

$$\begin{aligned}
| M(x, p, r) &- Ax + Ax/p + \cdots + (-1)^{\nu-1}Ax/p^{\nu} \\
&+ (-1)^{\nu}M(x/p^{\nu+1}, p, r-\nu-1) | < \tfrac{1}{2}\delta x(1 + 1/p + \cdots + 1/p^{\nu}) \\
&\leq \tfrac{1}{2}\delta x \sum_{s=0}^{\infty} 1/2^s = \delta x \text{ if } x/p^{\nu} \geq \Delta', \text{ that is if } x \geq \Delta'^2.
\end{aligned} \tag{4}$$

From (4) we get

$$\begin{aligned}
| M(x, p, r) - Ax \sum_{s=0}^{\infty}(-1)^s/p^s | &< \delta x + Ax\{1/p^{\nu+1} + 1/p^{\nu+2} + \cdots\} + \sqrt{x} \\
&\leqq \delta x + Ax/p^{\nu+1}(1 + \tfrac{1}{2} + \cdots) + \sqrt{x} \\
&\leqq \delta x + 2A\sqrt{x} + \sqrt{x} \text{ if } x \geq \Delta'^2.
\end{aligned}$$

Hence,

$$| M(x, p, r) - Ah(p)x | < \delta 2x \text{ if } x \geq \Delta = \Delta(\delta, e). \tag{5}$$

Suppose, now, that (1) is true for $d = d_j = \prod_{i=1}^{j} p_i$, $p_1, p_2, \cdots, p_j$ being different primes. We shall show that (1) is true for $d_{j+1} = \prod_{i=1}^{j+1} p_i$ also. We have evidently,

$$\begin{aligned}
M(x, d_{j+1}, r) &= M(x, d_j, r) \\
&\quad - M(x/p_{j+1}, d_j, r-1) + M(x/p_{j+1}{}^2, d_j, r-2) - \cdots.
\end{aligned}$$

Moreover

(15) $$r2^a \equiv r2^b \pmod{k} \text{ if and only if } a \equiv b \pmod{e'}.$$

Proof. Trivial.

LEMMA 5. *Let k be odd, r fixed and $\neq 0$; $(r,k) = t$.* Let $n = m^k qu \leq x$ *in the sense of* (10). *Suppose $f(q) \equiv \lambda' \pmod{e'}$. Then $n \subset N(k.r,x)$ if and only if*

(16) $$d(u) \equiv r2^{e'-\lambda'} \pmod{k}.$$

Proof. It is easily seen that $d(n) \equiv d(qu) \pmod{k}$ whether m is prime to qu or not. Hence, since $(q,u) = 1$,

$$d(n) \equiv d(q)d(u) \pmod{k}.$$

It follows that $n \subset N(k,r,x)$ if and only if

(17) $$d(q)d(u) \equiv r \pmod{k}.$$

Let $f(q) \equiv \lambda \pmod{e}$; then

(18) $$d(q) = 2^{f(q)} \equiv 2^\lambda \pmod{k}.$$

If now $n \subset N(k,r,x)$, then it follows from (17) that $d(u) \equiv r2^{e-\lambda} \pmod{k}$. Hence from (15), $d(u) \equiv r2^{e'-\lambda'} \pmod{k}$. Conversely, if $d(u) \equiv r2^{e'-\lambda'} \pmod{k}$, then it follows from (18) and (15) that $d(q)d(u) \equiv r2^{e'-\lambda'+\lambda} \equiv r \pmod{k}$, since $e' - \lambda' + \lambda \equiv 0 \pmod{e'}$.

Remark. If $d(u) \equiv r2^a \pmod{k}$ and u is fixed then integers $n = m^k qu$ belong to $N(k,r,x)$ (for different values of m and q) if and only if

(19) $$f(q) \equiv e' - \alpha \pmod{e'} \text{ and } (q,u) = 1.$$

LEMMA 6. *If k is odd and $n = m^k qu \leq x$ then $n \subset N(k,0,x)$ if, and only if, $d(u) \equiv o \pmod{k}$.*

Proof. This follows immediately from (17) which is true even if $r = 0$.

3. Proof of the main theorem.

THEOREM 1. *If k is odd, $1 \leq r \leq k-1$, then $N(k,r,x) \sim Bx$.*

Proof. If $n = m^k qu \leq x$ where $d(u) \not\equiv r2^a \pmod{k}$ then by Lemma 5 n does not belong to $N(k,r,x)$. Denote by u' an integer of the set u such that

(20) $$d(u') \equiv r2^a \pmod{k},$$

where we may regard $1 \leq \alpha \leq e'$ by virtue of Lemma 4. From (19) we see that the contribution to $N(k,r,x)$ *by a fixed* u' is equal to the number of integers

$$(21)\qquad m^k q u' \leq x \text{ where } f(q) \equiv e' - \alpha \pmod{e'} \text{ and } (q,u') = 1.$$

This, in the notation of Lemma 1, is equal to

$$(22)\qquad \sum_{m^k \leq x/u'} M(x/m^k u', \prod_{p|u'} p, e' - \alpha, e') = \sum_{m^k \leq x^{7/8}/u'} + \sum_{x^{7/8}/u' < m^k \leq x/u'}$$
$$= S_1 + S_2, \text{ say.}$$

$S_2 = O(x^{1/8+1/k}) = O(x^{11/24})$ since $k \geq 3$. Suppose that the fixed number u' is not greater than $x^{3/4}$. Then from Lemma 1, since $h(\prod_{p|u'} p) = h(u')$ and $x/m^k u' > x^{1/8}$, there exists, for every positive number δ, a $\Delta = \Delta(\delta, e')$ independent of u' such that

$$| S_1 - \sum_{m^k \leq x^{7/8}/u'} xh(u')/e'\zeta(2) m^k u' | < \delta 2^{f(u')} x/u' \sum_{m=1} 1/m^k \text{ if } x \geq \Delta^8.$$

Hence, since $h(u') \leq 1$,

$$| S_1 - xh(u')\zeta(k)/e'\zeta(2)u' | < \delta\zeta(k) 2^{f(u')} x/u' + x/e'\zeta(2)u' \sum_{m > x^{1/8k}}^{\infty} 1/m^k$$
$$\text{if } x \geq \Delta^8$$
$$< \delta(2^{f(u')}/u')x\zeta(k) + \delta x\zeta(k)/u' \text{ if } x \geq \Delta_0$$

Δ_0 being independent of u'. Hence,

$S_1 = xh(u')\zeta(k)/e'\zeta(2)u' + O(\delta 2^{f(u')}/u')x)$ for every arbitrarily small δ where the constant in O is independent of δ and u'. From this and from (22) the contribution to (21) by every fixed $u' \leq x^{3/4}$ is equal to

$$(23)\qquad xh(u')\zeta(k)/e'\zeta(2)u' + O(\delta x 2^{f(u')}/u') + O(x^{11/24}).$$

Now the total contribution to (21) by all $u' > x^{3/4}$ is $O(x^{3/4})$ since the contribution by every such u' is at most equal to $x^{1/4}$ and the number of $u' \leq x$ is $O(x^{1/2})$ by Lemma 2. From this and from (19) we get,

$$N(k,r,x) = x\zeta(k)/e'\zeta(2) \sum_{u' \leq x^{3/4}} h(u')/u'$$
$$+ O(\delta x \sum_{u' \leq x^{3/4}} 2^{f(u')}/u') + O(x^{11/24} \sum_{u' \leq x^{3/4}}) + O(x^{3/4}).$$

The series $\sum_{u'} h(u')/u'$ and $\sum_{u'} 2^{f(u')}/u'$ are both convergent; for if $R(x)$ be the number of $u \leq x$ then,

$$\sum_{u'} 1/u'^a \leq \sum_{u} 1/u^a = O \sum_{n=1}^{\infty} \{R(n) - R(n-1)\}/n^a = O \sum_{n=1}^{\infty} R(n)/n^{1+a}$$
$$= o \sum_{n=1}^{\infty} 1/n^{a+1/2} \text{ by Lemma 2.}$$

Hence $\Sigma 1/u'^{a}$ converges if $a > 1/2$. From this the convergence of the series in question follows since $h(u') \leq 1$ and $2^{f(u')} \leq d(u') = O(u'^{\epsilon})$ for every $\epsilon > 0$. Finally,

$$N(k, r, x) = x\zeta(k)/e'\zeta(2) \sum_{u'} h(u')/u' + O(\delta x) + o(x)$$
$$\sim Bx \text{ for every positive } \delta$$

(since the constant in $O(\delta x)$ is independent of δ).where,

$$(24) \qquad B = C\zeta(k)/e'\zeta(2) \quad \text{and} \quad C = \sum_{u'} h(u')/u',$$

u' running over all the integers u satisfying (20). To show that $B > 0$ we have only to prove that $C > 0$; that is that the set u' is not empty. Evidently if p is any prime, then p^{r-1} is a u' if $k - 1 \geq r \geq 3$. If $r = 1$ or 2 we see that 1 is a u'. For $d(1) \equiv 1 \pmod{k}$ and $d(1) \equiv 2\, 2^{e'-1} \equiv 1 \pmod{k}$. Hence (20) is satisfied for $u' = 1$.

THEOREM 2. *If* $k = 2^{l}k'$; $(2, k') = 1$

$$r = 2^{t}m'; \ (2, m') = 1; \ 0 \leq t \leq l - 1$$

then $N(k, r, x) = o(x)$.

Proof. If $n = m^{k}qu \leq x$ then, as in (17), n belongs to $N(k, r, x)$ if and only if $d(q)d(u) \equiv 2^{t}m' \pmod{2^{l}k'}$. Dividing this congruence by 2^{t} we see that it has no solution with $f(q) \geq l$ if $t \leq l - 1$. The result, therefore, follows from Lemma 3.

THEOREM 3. $N(2^{l}k', 2^{l}m', x) \sim Bx$, $B > 0$, *where* k' *is odd and* $1 \leq m' \leq k' - 1$.

Proof. Let

$$(25) \qquad 2^{l}m' \equiv m'' \pmod{k'}.$$

We note in the first instance that if

$$(26) \qquad n \subset N(2^{l}k', 2^{l}m', x) \quad \text{then} \quad n \subset N(k', m'', x).$$

Also if $n \subset N(k', m'', x)$ then $d(n) \equiv m'' \pmod{k'}$. Hence from (25), $d(n) \equiv zk' + 2^{l}m' \pmod{2^{l}k'}$ where $z = 0, 1, \cdots$ or $2^{l} - 1$. Hence either (27) $d(n) \equiv 2^{l}m' \pmod{2^{l}k'}$ or $d(n) \equiv s \pmod{2^{l}k'}$ where $2^{l} \nmid s$.

From (26) and (27) it follows that

$$N(k', m'', x) = N(2^{l}k', 2^{l}m', x) + X$$

where $X = O \sum_{s} N(2^{l}k', s, x)$, s running through all the residue classes $\pmod{2^{l}k'}$ for which $2^{l} \nmid s$. Hence $X = o(x)$ by Theorem 2. Therefore,

$$N(2^l k', 2^l m', x) \sim N(k', m'', x) \sim Bx \tag{28}$$

by Theorem 1, where $B > 0$.

THEOREM 4. $N(k, 0, x) \sim Bx$.

Proof. It is sufficient to consider the case when k is odd, since it *then* follows as in (28) that

$$N(2^l k', 0, x) \sim N(k', 0, x) \text{ if } k' \text{ is odd.} \tag{28'}$$

It follows from Lemma 6 that integers n belonging to $N(k, 0, x)$ are of the form

$$m^k q u'' \leq x \text{ where } (q, u'') = 1, \tag{29}$$

q being any square free integer, u'' being a u for which $d(u) \equiv 0 \pmod{k}$. If now $Q(x, d)$ denotes the number of square free integers $q \leq x$ for which $(q, d) = 1$ then, from Lemma 1

$$Q(x, d) = M(x, d, 1) + \cdots + M(x, d, e) = h(d)x/\zeta(2) + O(\delta 2^{f(d)} x) \tag{30}$$

for every $\delta > 0$, the constant in $O(\delta 2^{f(d)} x)$ being independent of δ and d. Also

$$N(k, 0, x) = \sum_{m^k u'' \leq x} Q(x/m^k u'', \prod_{p | u''} p)$$

the summation being with respect to both m and u''. Using (30) we get as in Theorem 1

$$N(k, r, x) \sim Bx$$

where

$$B = \zeta(k)/\zeta(2) \sum_{u''} h(u'')/u'', \tag{31}$$

where u'' runs over all the integers of the set u for which $d(u) \equiv 0 \pmod{k}$. To prove that $B > 0$ we have to show that the set u'' is not empty. Evidently p^{k-1} is a u'' if p is any prime.

The proof of the main theorem now follows from Theorems 1 to 4.

4. Miscellaneous results.

THEOREM 5. *Let k be odd; $(r, k) = t$ and, as in Lemma 4, let e' be the exponent of* 2 (mod k/t). *Then,*

$$N(2^l k, 2^l r, x) \sim N(k, r 2^\beta, x); \qquad 1 \leq \beta \leq e'. \tag{32}$$

In particular when $l = 0$ and 2 is the primitive root of k, we get $N(k, r, x) \sim N(k, r', x)$ if neither r nor r' is zero.

Proof. Let k be odd; $1 \leq r \leq k-1$. We note in the first instance that only those u' which satisfy (20) contribute to $N(k, r2^\beta, x)$. Also (23) still remains true since the dominant term of (22) is independent of α. Hence we see that (24) is true for $r2^\beta$ also. Hence, (32) is true for odd values of k and all values of r, since when $r = 0$ it reduces to an identity.

If now $k = 2^l$, $(k', 2) = 1$ then we get from (28) that

$$N(2^l k', 2^l m', x) \sim N(k', m'', x) \tag{33}$$

where $m'' = 2^l m' \pmod{k'}$. But by the case just considered,

$$N(k', m'', x) \sim N(k', m'' 2^\gamma, x) \text{ where } 1 \leq \gamma \leq e'.$$

Also $m'2^\beta$ and $m''2^\gamma$, where $1 \leq \beta \leq e'$; $1 \leq \gamma \leq e'$, run over the same classes of residues (mod k') by virtue of (25). Hence,

$$N(2^l k', 2^l m', x) \sim N(k', m' 2^\beta, x) \text{ where } 1 \leq \beta \leq e'.$$

Writing k for k' and r for m' we get (32) for $1 \leq r \leq k-1$. If $r = 0$ then (32) reduces to (28′). Hence the theorem is proved.

THEOREM 6.

$$N(2^l k, 0, x) \sim \{1 - \zeta(k)/\zeta(k-1)\}x, \textit{ if } k \textit{ is a prime.} \tag{34}$$

Proof. By virtue of (28′) we may suppose $l = 0$ without loss of generality. To prove the result we must evaluate the sum of the series $\sum_{u''} h(u'')/u''$ where u'' runs over those u for which $d(u) \equiv 0 \pmod{k}$. Since k is a prime, every u'' is of the form

$$q_2{}^2 \cdots k_{k-2}{}^{k-2} q'_{k-1}{}^{k-1}$$

where $q_2, \cdots, q_{k-2}$ run over all square free numbers and q'_{k-1} runs over all square free numbers except unity. $q_2, \cdots, q'_{k-1}$ in every such number are relatively prime. Hence,

$$\sum_{u''} h(u'')/u'' \\ = \sum_{q_2, \cdots, q'_{k-1}} (1/q'_{k-2}{}^{k-2}) \Big(\prod_{p \mid q'_{k-1}} 1/(p+1) \cdots (1/q_2) \Big(\prod_{p \mid q_2} 1/(p+1) \Big) \Big).$$

In every term of this multiple series $q_2, \cdots, q'_{k-1}$ are mutually prime. Hence fixing $q_3, \cdots, q'_{k-1}$ and summing first with respect to q_2 we get

$$\begin{aligned} \sum (1/q_2) \Big(\prod_{p \mid q_2} 1/(p+1) \Big) &= \prod_p \{1 + 1/(p+1)\} / \prod_{p \mid q_3 \cdots q'_{k-1}} \{1 + 1/p(p+1)\} \\ &= \{\prod_p (1 - 1/p^3)\} \prod_{p \mid q_3 \cdots q'_{k-1}} \{p(p+1)/(p^2+p+1)\} / \prod_p (1 - 1/p^2) \\ &= \zeta(2)/\zeta(3) \prod_{p \mid q_3 \cdots q'_{k-1}} \{p(p+1)/(p^2+p+1)\}. \end{aligned}$$

Hence the series summation with respect to q_2 reduces to

$$\zeta(2)/\zeta(3)\sum(1/{q'_{k-2}}^{k-3})\Big(\prod_{p|q'_{k-1}}1/(p^2+p+1)\Big)\cdots(1/q_3)\Big(\prod_{p|q_3}1/(p^2+p+1)\Big).$$

We now fix $q_4, \cdots, q'_{k-1}$ and sum with respect to q_3 and so on. Finally, noting that q'_{k-1} does not take the value 1 we get

$$(35)\quad \sum h(u'')/u'' = \frac{\zeta(2)}{\zeta(3)}\cdot\frac{\zeta(3)}{\zeta(4)}\cdots\frac{\zeta(k-2)}{\zeta(k-1)}\left\{\frac{\zeta(k-1)}{\zeta(k)}-1\right\}$$
$$= \frac{\zeta(2)}{\zeta(k-1)}\left\{\frac{\zeta(k-1)}{\zeta(k)}-1\right\}.$$

From (35) and (31) the theorem follows.

Note. Let 2^l be the highest power of 2 that divides k and $1 \leq t \leq l-1$. Then, by sharpening Lemma 4 we can prove that

$$N(k,r,x) \sim D(x/\log x)(\log\log x)^{t-1} \text{ if } 2^t \| r$$

where $D = D(k,r)$ is a constant > 0. Further if r is odd then it is easy to see that $N(k,r,x) = O(x^{1/2})$, *if k is even.*

Calcutta University,
Calcutta, India.

REFERENCES.

[1] S. S. Pillai, *Journal of the Indian Mathematical Society,* vol. 6, N. S. (1942).
[2] S. S. Pillai, *Proceedings of the Indian Academy of Sciences,* vol. 11 (1940).
[3] Sigmund Selberg, *Mathematische Zeitschrift,* Band 44 (1938), pp. 306-318.

THE RECTILINEAR MOTION OF A GAS II.*

By Monroe H. Martin.

In the rectilinear motion of a gas we face the problem of determining the velocity u, the density ρ and the pressure p as a function of the position x and time t. In addition to meeting any initial conditions, which may be imposed on the motion of the gas, these three functions must satisfy the underdetermined system of partial differential equations

$$(u_t + u_x u)\rho + p_x = 0, \qquad \rho_t + (\rho u)_x = 0. \tag{1}$$

If the second equation in (1) is multiplied by u and added to the first, the system (1) is seen to be equivalent to the system

$$(\rho u)_t + (p + u^2\rho)_x = 0, \qquad \rho_t + (u\rho)_x = 0. \tag{2}$$

Assuming that a solution u, ρ, p of this system is at hand, there exist functions v, z of x, t such that the system of Pfaff

$$dv = -u\rho dx + (p + u^2\rho)\,dt, \qquad dz = \rho dx - u\rho dt,$$

possesses a solution in which the variables u, ρ, p, z, v are all functions of x, t. Conversely, given a solution of the system of Pfaff in which u, ρ, p, z, v are functions of x, t, one has in u, ρ, p a solution of (2), and hence a solution of (1).

The usual procedure is to make (1) a determined system by adding the characteristic equation $p = p(\rho)$ of the gas. The question has been raised,[1] however, whether the characteristic equation, known to hold for a perfect gas at rest, will continue to hold for a gas in motion. In this note, instead of assuming a characteristic equation for the gas, we supplement (1) by a third partial differential equation for u, p. This partial differential equation represents the necessary and sufficient condition that the Principle of Conservation of Energy hold for a gas in motion.

According to thermodynamical principles[2] the internal energy e per unit mass of a perfect gas is given by $e = p/(\gamma - 1)\rho$, where γ denotes the ratio of the two specific heats of the gas. The total energy E of the gas in a right

---

* Received September 9, 1944.

[1] H. Weber, "*Die partiellen Differentialgleichungen der Mathematischen Physik,*" Braunschweig, vol. 2 (1919), p. 553.

[2] H. Weber, *op. cit.*, p. 507.

circular cylindrical tube of unit cross-sectional area, with bases on the planes $x = x_1$, $x = x_2$, is accordingly

$$E = \int_{x_1}^{x_2} (\tfrac{1}{2}u^2 + e)\rho\, dx, \qquad x_1 < x_2.$$

The Principle of Conservation of Energy holds if, and only if, the rate at which energy is being lost from the tube equals the rate at which the two bases of the tube do work on the surrounding gas,[3] that is

$$\dot{E} = p_1u_1 - p_2u_2 = -\int_{x_1}^{x_2} (pu)_x\, dx,$$

where here and elsewhere the subscripts 1, 2 on a variable indicate the values of this variable on the planes $x = x_1$, $x = x_2$. Using the above expression for E, it may be shown that this equation holds in conjunction with (1) if, and only if, u, p satisfy the partial differential equation

$$p_t + up_x + \gamma p u_x = 0.$$

If we set $p = \sigma^\gamma$, this equation becomes $\sigma_t + (\sigma u)_x = 0$, analogous to the equation of continuity in (1), and we have a determined system of partial differential equations

$$(u_t + u_x u)\rho + p_x = 0, \qquad \rho_t + (u\rho)_x = 0, \qquad \sigma_t + (u\sigma)_x = 0,$$

for the unknown functions u, ρ, σ. As above, we see that the problem of finding a solution $u, \rho, p = \sigma^\gamma$ of this system is equivalent to the problem of finding functions u, ρ, σ, v, z, w of x, t satisfying the system of Pfaff

$$dv = -u\rho\, dx + (\sigma^\gamma + u^2\rho)\, dt, \quad dz = \rho\, dx - u\rho\, dt, \quad dw = \sigma\, dx - u\sigma\, dt.$$

It will be observed that $\sigma dz - \rho dw = 0$, so that the functions z, w of x, t are always functionally dependent, $w = w(z)$. Now $\sigma = w'\rho$, where w' denotes the derivative of w with respect to z. Thus we have $p = (w'\rho)^\gamma$ and the above system of Pfaff may be replaced by a system of two equations of Pfaff in six variables x, t, u, ρ, z, v,

$$dv = -u\,dz + (w'\rho)^\gamma dt, \qquad dz = \rho\, dx - u\rho\, dt, \tag{3}$$

containing an arbitrary function $w' = w'(z)$.

The difference $z_2 - z_1$ of z for fixed t yields the mass of the gas in the cylindrical tube at the instant t while $v_1 - v_2$ yields its momentum. Fixing our attention on a particular gas particle, the value of z assigned to this gas particle remains constant throughout the motion of the gas. *The trajectories of gas particles in the* (x, t) *plane are accordingly represented by the one parameter family of curves* $z = const.$

---

[3] Naturally this involves the hypothesis that there is no loss of energy by heat conduction or radiation.

Since z is constant along a trajectory, the relation between p and ρ along a trajectory follows the adiabatic law $p = k^2\rho^\gamma$, with $k^2 = w'^\gamma$ varying from trajectory to trajectory unless $w' = \text{const.}$ From thermodynamical considerations it is known[2] that $S = c_v \log k^2$, where S denotes the entropy and c_v the specific heat of the gas at constant volume. Hence $S = \gamma c_v \log w'$ and if we assume $c_v = \text{constant}$ throughout the gas, we have the following result. *The entropy remains constant along a trajectory and varies from trajectory to trajectory with* $\log w'$.

We observe that the first equation in (3) is of class 5 and assumes normal form if we set $p = (w'\rho)^\gamma$, $q = -u$. Accordingly (3) is equivalent to the system of Pfaff

$$(4) \qquad dv = pdt + qdz, \qquad dx = -qdt + w'p^n dz, \qquad n = -1/\gamma,$$

in the six variables t, z, v, p, q, x.

Given functions p, q, v, x of z, t satisfying (4), we have $p = v_t$, $q = v_z$ and therefore, from the second equation of (4), v is a solution of the partial differential equation

$$(5) \qquad v_{zz} + nw'v_t^{n-1}v_{tt} = 0.$$

Conversely starting with a solution $v = v(z, t)$ of (5), we find u, ρ, p expressed in terms of z, t by

$$(6) \qquad u = -v_z, \qquad \rho = 1/w'v_t^n, \qquad p = v_t.$$

To express u, ρ, p in terms of x, t, one solves

$$x = x(z, t) = \int \{-v_z dt + w'v_t^n dz\},$$

for z as a function of x, t.

If we take u, ρ, p as known functions of x for $t = 0$ and assume $\rho > 0$ everywhere, $z = \int \rho dx$ is a monotonic increasing function of x. Accordingly we may regard u, ρ, p as known functions of z for $t = 0$. From (6) it appears that not only are the values of v_t, v_z prescribed for $t = 0$, but also the function $w'(z) > 0$ appearing in (5) is uniquely determined. *The problem of determining the rectilinear motion of a gas subsequent to prescribed initial conditions on its velocity, density, and pressure is accordingly reduced to the solution of the Problem of Cauchy for the partial differential equation* (5), *the function* $w' = w'(z)$ *having previously been determined from the initial conditions.*

Once w' is determined, solutions of (5) may be obtained by the usual process of separation of variables. Thus $v = TZ$ is a solution, provided $T = T(t)$, $Z = Z(z)$ are solutions of

$$Z'' - knw'Z^n = 0, \qquad T'^{n-1}T'' + kT = 0, \qquad k = \text{const.}$$

In the special case where the entropy is constant throughout the gas ($w' = \text{const.}$), Legendre's transformation transforms (5) into a linear partial differential equation of second order and hyperbolic type. After reduction to normal form, Riemann's method may be applied to obtain a solution for the Problem of Cauchy set by the initial conditions.[4]

It may be of some interest to point out that in the general case Ampère's contact transformation

$$t = -P, \qquad z = Z, \qquad v = V - PT, \qquad p = T, \qquad q = Q,$$

carries (5) into a Monge-Ampère partial differential equation

$$V_{TT}V_{ZZ} - V^2_{ZT} + \lambda^2 = 0, \qquad \lambda^2 = -nT^{n-1}w'.$$

The characteristics of this partial differential equation comprise the asymptotic lines on the solution surfaces.[5]

UNIVERSITY OF MARYLAND,
COLLEGE PARK, MD.

---

[4] See B. Riemann, "Über die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite," *Werke*, Leipzig (1892), 2nd Edition, pp. 156-175. Further references are given in the paper by the author, "The rectilinear motion of a gas," *American Journal of Mathematics*, vol. 65 (1943), pp. 391-407.

[5] E. Goursat, *Leçons sur l'integration des équations aux dérivées partielles du second ordre*, vol. 1, Paris (1896), pp. 147-148.

ON THE ASYMPTOTIC BEHAVIOR OF THE RIEMANN ξ-FUNCTION.*

By E. K. Haviland.

As is well known, the function, $\xi(s)$, defined by

$$\xi(s) = \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s),$$

possesses, as its only zeros, the non-trivial zeros of the Riemann ζ-function.[1] Furthermore, if $M(r)$ denotes the maximum of $|\xi(\frac{1}{2}+is)|$ on the circle $|s| = r$, then

$$\log M(r) \sim \tfrac{1}{2}r\log r. \tag{1}$$

This asymptotic relation is a well-known result of Hadamard and is more than sufficient for his purpose in applying his theory of entire functions to the Riemann ξ-function.[2] Actually, he defines $M(r)$ to be the maximum of $|\xi(s)|$ on the circle $|s| = r$, which does not affect the formula (1), but would affect the more delicate relations to be considered below.

There arises the desirability of an asymptotic formula for $M(r)$ itself, since the situation, as observed by Wintner, is as follows: Every *individual* zero of ξ has a number-theoretical significance (the explicit meaning of which is unknown). On the other hand, it is clear from the Weierstrass-Hadamard factorization

$$\xi(\tfrac{1}{2}+is) = \prod_{\rho}(1 - s^2/\rho^2) \tag{2}$$

that the omission or the "wrong" position (i. e., not in accordance with Riemann's hypothesis) of a finite number, or even of a sufficiently "thin" infinite sequence, of zeros has no influence whatsoever on (1), while each of these individual zeros will make a contribution which appears in the asymptotic behavior of $M(r)$, if this behavior is not weakened to standards of the logarithmic scale.

In fact, it turns out that the true law to which the classical result (1) is the logarithmic approximation is

---

* Received April 10, 1945.

[1] Cf., e. g., A. E. Ingham, *op. cit.*, p. 48.

[2] *Ibid.*, pp. 56-58.

$$M(r) \sim (\tfrac{1}{2}\pi)^{1/4}(2\pi e)^{-\frac{1}{2}r} r^{\frac{1}{2}r+7/4}, \quad \text{as} \quad r \rightarrow \infty. \tag{3}$$

But it is clear from (2) that the contribution of each of the factors to $M(r)$ is $r^2/|\rho|^2$, so that the contribution of each of the zeros to the right-hand side of (3) is explicit indeed. Incidentally, the 7/4 and the other elements occurring in (3) may be thought of as manifestations of a Fourier transformation of the corresponding elements in the Riemann–v. Mangoldt formula. In this connection, see the remarks of Wintner.[3]

Actually, (3) can be improved by showing the existence of an infinite asymptotic series

$$M(r) \sim (\tfrac{1}{2}\pi)^{1/4}(2\pi e)^{-\frac{1}{2}r} r^{\frac{1}{2}r+7/4} \sum_{n=0}^{\infty} C_n/r^n, \tag{4}$$

where the C's are numerical coefficients and (3) means merely that the numerical value of C_0 is 1.

For the purposes of carrying out the proof of these results, it is more convenient to deal, not with $\xi(s)$ as above defined, but with

$$\Xi(t) = \xi(\tfrac{1}{2} + it), \qquad t \text{ complex},$$

which is $\xi(t)$ in the notation of Riemann. For $\Xi(t)$, one has the formula:[4]

$$\Xi(t) = 2\int_0^{\infty} \Phi(u) \cos 2ut\, du,$$

where

$$\Phi(u) = 4\sum_{n=1}^{\infty} (2n^4\pi^2 e^{9u} - 3n^2\pi e^{5u}) \exp[-n^2\pi e^{4u}],$$

and it is known that $\Phi(u)$ is an even function of u, positive for all real values of u and such that $\Phi'(u) < 0$ for every $u > 0$.[5]

Corresponding to this integral representation of $\Xi(t)$, it is convenient to re-define $M(r)$ so as to denote

$$\begin{aligned} M(r) &= \operatorname*{Max}_{|t| \leqq r} |\Xi(t)| = \Xi(ir) = 2\int_0^{\infty} \Phi(u) \cosh 2ur\, du \\ &= \int_0^{\infty} \Phi(u) e^{2ru} du + \int_0^{\infty} \Phi(u) e^{-2ru} du = I + II. \end{aligned} \tag{5}$$

and our problem is to examine the behavior of $M(r)$, as $r \rightarrow \infty$.

---

[3] A. Wintner, *loc. cit.* 5, p. 58.

[4] Cf., e. g., E. C. Titchmarsh, *op. cit.*, p. 45.

[5] Cf. A. Wintner, *loc. cit.* 4, p. 82, and further references there given.

We begin by writing

$$\begin{aligned}(6)\qquad \Phi(u) &= 4\sum_{n=1}^{k}(2n^4\pi^2e^{9u}-3n^2\pi e^{5u})\exp[-n^2\pi e^{4u}]\\ &\quad+4\sum_{n=k+1}^{\infty}(2n^4\pi^2e^{9u}-3n^2\pi e^{5u})\exp[-n^2\pi e^{4u}]\\ &=\Phi_k(u)+\Psi_k(u),\end{aligned}$$

where $0 \leqq u < +\infty$. In the foregoing infinite series, we replace e^u by x, so that it becomes

$$4\sum_{n=k+1}^{\infty}(2n^4\pi^2x^9-3n^2\pi x^5)\exp[-n^2\pi x^4],\qquad 1\leqq x<+\infty.$$

Since $2n^2\pi > 3$, $n \geqq 1$, it follows that

(i) each term is positive and hence the series, if convergent, is positive for all x in the range indicated;

(ii) the series is term for term less than

$$8x\sum_{n=k+1}^{\infty}n^4(\pi x^4)^2\exp[-n^2\pi x^4].$$

Under the substitution $\pi x^4 = y$, this last series becomes

$$(7)\qquad 8(y/\pi)^{1/4}\sum_{n=k+1}^{\infty}n^4y^2\exp[-n^2y],\qquad \pi\leqq y<+\infty.,$$

Since $d(x^2e^{-x})/dx = xe^{-x}(2-x) < 0$ for $x > 2$, it follows that, for fixed $y \geqq \pi$ and for $n \geqq 2$, the general term of (7) is a positive monotone decreasing function of n, so that

$$\sum_{n=k+1}^{\infty}n^4y^2\exp[-n^2y]<(k+1)^4y^2\exp[-(k+1)^2y]+\int_{k+1}^{\infty}y^2s^4\exp[-ys^2]ds,$$

$k \geqq 1$. This last integral is less than

$$\int_{k+1}^{\infty}y^2s^5\exp[-ys^2]ds=1/2y\int_{y(k+1)^2}^{\infty}x^2e^{-x}dx,\quad\text{where}\quad x=ys^2.$$

Partial integration then shows that

$$\begin{aligned}\sum_{n=k+1}^{\infty}n^4y^2\exp[-n^2y]&<(k+1)^4y^2\exp[-(k+1)^2y]\\ &\quad+(2y)^{-1}\exp[-(k+1)^2y]\{(k+1)^4y^2+2(k+1)^2y+2\}\\ &<2(k+1)^4y^2\exp[-(k+1)^2y],\end{aligned}$$

in view of $k \geq 1$, $y \geq \pi$. Consequently,

$$(8) \quad 0 < \Psi_k(u) \equiv \Psi_k(u(y)) < 16(y/\pi)^{1/4}(k+1)^4 y^2 \exp[-(k+1)^2 y],$$

and so

$$\begin{aligned}\int_0^\infty \Psi_k(u) e^{2ru} du &= \tfrac{1}{4}\pi^{-\frac{1}{2}r} \int_\pi^\infty \Psi_k(u(y)) y^{\frac{1}{2}r-1} dy \\ &< 4(k+1)^4 \pi^{-\frac{1}{2}r-1/4} \int_\pi^\infty y^{\frac{1}{2}r+5/4} \exp[-(k+1)^2 y] dy \\ &= 4[\pi(k+1)^2]^{-\frac{1}{2}r-1/4} \int_{(k+1)^2\pi}^\infty t^{\frac{1}{2}r+5/4} e^{-t} dt \\ &< 4[\pi(k+1)^2]^{-\frac{1}{2}r-1/4} \Gamma(\tfrac{1}{2}r + 9/4),\end{aligned}$$

where $(k+1)^2 y = t$.

With the help of the preceding inequalities, it may now readily be shown[6] that $\int_0^\infty \Phi(u) e^{2ru} du$ may, for fixed r, be integrated term by term. Moreover, as we have just seen,

$$(9) \quad \int_0^\infty \Psi_k(u) e^{2ru} du < [\pi(k+1)^2]^{-\frac{1}{2}r-1/4}(r+5/2)(r+\tfrac{1}{2})\Gamma(\tfrac{1}{2}r+1/4).$$

In the finite set of terms obtained from $\int_0^\infty \Phi_k(u) e^{2ru} du$, that one will contribute least which corresponds to the term

$$-12k^2\pi e^{5u} \exp[-k^2\pi e^{4u}]$$

of $\Phi_k(u)$, for such a term is in absolute value less than the corresponding positive term in (6) and

$$d(x^2 \exp[-ax^2])/dx = 2x \exp[-ax^2](1 - ax^2) < 0,$$

if $a > 1$, $x > 1$, as is here the case, since $a = \pi e^{4u}$, $u \geq 0$ and $k \geq 1$. Then let

$$B_k(r) = -12k^2\pi \int_0^\infty e^{2ru+5u} \exp[-k^2\pi e^{4u}] du.$$

If we make the substitution $\pi e^{4u} = y$, we obtain

$$B_k(r) = -3k^2\pi^{-\frac{1}{2}r-1/4} \int_\pi^\infty y^{\frac{1}{2}r+1/4} \exp[-k^2 y] dy.$$

The further substitution $k^2 y = t$ leads to

---

[6] Cf., e. g., W. F. Osgood, *op. cit.*, p. 287.

$$B_k(r) = -3(\pi k^2)^{-\frac{1}{2}r-1/4} \int_{\pi k^2}^{\infty} t^{\frac{1}{2}r+1/4} e^{-t}\,dt$$

$$= -3(\pi k^2)^{-\frac{1}{2}r-1/4} \int_{0}^{\infty} t^{\frac{1}{2}r+1/4} e^{-t}\,dt + o(1),$$

since

$$0 < (\pi k^2)^{-\frac{1}{2}r-1/4} \int_{0}^{\pi k^2} t^{\frac{1}{2}r+1/4} e^{-t}\,dt < (\pi k^2)^{-\frac{1}{2}r-1/4} \cdot (\pi k^2)^{\frac{1}{2}r+5/4}/(\tfrac{1}{2}r + 5/4)$$

$$= \pi k^2/(\tfrac{1}{2}r + 5/4) = o(1), \quad \text{as} \quad r \to \infty.$$

Hence

$$B_k(r) = -3/2\,(\pi k^2)^{-\frac{1}{2}r-1/4}(r + \tfrac{1}{2})\Gamma(\tfrac{1}{2}r + 1/4) + o(1). \tag{10}$$

Comparison of (10) with (9) shows that

$$\int_0^{\infty} \Psi_k(u)\, e^{2ru}\,du = o(B_k(r)), \quad \text{as} \quad r \to \infty.$$

Furthermore, on setting $k = 1$ in (8) and observing that $y = \pi e^{4u}$, we find that

$$0 < \Psi_1(u) < 256\pi^2 e^{9u} \exp\,[-4\pi e^{4u}].$$

This, together with (6) and (5), shows that II is $O(1)$, hence $o(B_k(r))$, for every fixed k, as $r \to \infty$.

Consequently, it is possible to obtain an expression for $M(r)$ in terms of an infinite asymptotic series. In carrying out this process, we first calculate

$$A_k(r) = 8k^4\pi^2 \int_0^{\infty} e^{2ru+9u} \exp\,[-k^2\pi e^{4u}]\,du.$$

The calculation, which is exactly similar to that for $B_k(r)$ above, shows that

$$A_k(r) = \tfrac{1}{2}(\pi k^2)^{-\frac{1}{2}r-1/4}(r + 5/2)(r + \tfrac{1}{2})\Gamma(\tfrac{1}{2}r + 1/4) + o(1), \quad \text{as} \quad r \to \infty.$$

Hence, if $k = 1$,

$$\int_0^{\infty} \Phi_1(u)\, e^{2ru}\,du = A_1(r) + B_1(r)$$

$$= \tfrac{1}{2}\pi^{-\frac{1}{2}r-1/4}(r + 5/2)(r + \tfrac{1}{2})\Gamma(\tfrac{1}{2}r + 1/4)$$

$$- (3/2)\pi^{-\frac{1}{2}r-1/4}(r + \tfrac{1}{2})\Gamma(\tfrac{1}{2}r + 1/4) + o(1)$$

$$= \tfrac{1}{2}\pi^{-\frac{1}{2}r-1/4}(r + \tfrac{1}{2})\Gamma(\tfrac{1}{2}r + 1/4)(r - \tfrac{1}{2}) + o(1)$$

$$= \tfrac{1}{2}\pi^{-\frac{1}{2}r-1/4}(r^2 - 1/4)\Gamma(\tfrac{1}{2}r + 1/4) + o(1). \tag{11}$$

But $\Gamma(x) = (2\pi)^{\frac{1}{2}} x^{x-\frac{1}{2}} e^{-x+\mu(x)}$, where $\mu(x) = (12x)^{-1} + O(x^{-2})$, so that a straightforward calculation shows

$$\Gamma(\tfrac{1}{2}r + 1/4) = (2\pi)^{\frac{1}{2}}(\tfrac{1}{2}r)^{\frac{1}{2}r-1/4}e^{-\frac{1}{2}r}(1 - 1/(48r) + O(r^{-2}))$$

Substitution of this result into (11) gives

$$\begin{aligned} I_1 &= \int_0^\infty \Phi_1(u)e^{2ru}du \\ &= \tfrac{1}{2}\pi^{-\frac{1}{2}r-1/4}(r^2 - 1/4)(2\pi)^{\frac{1}{2}}(\tfrac{1}{2}r)^{\frac{1}{2}r}(2/r)^{1/4}e^{-\frac{1}{2}r}(1 - 1/(48r) + O(r^{-2})) \\ &= (\pi/2)^{1/4}r^{\frac{1}{2}r}(2\pi e)^{-\frac{1}{2}r}(r^{7/4} - (4r^{1/4})^{-1}(1 - 1/(48r) + O(r^{-2})) \\ &= (\pi/2)^{1/4}r^{\frac{1}{2}r}(2\pi e)^{-\frac{1}{2}r}(r^{7/4} - r^{3/4}/48 + O(r^{-1/4})), \end{aligned}$$

and the contributions of $\int_0^\infty \Psi_1(u)e^{2ru}du$, II may be merged into the O-term, so that the expansion calculated for I_1 serves also for $M(r)$. (Cf. (5), (6)). From this it is seen that the appraisal of $\int_0^\infty \Psi_{k_0}(u)e^{2ru}du$ could have been confined to the case $k = 1$.

THE LINCOLN UNIVERSITY,
CHESTER COUNTY, PENNSYLVANIA.

REFERENCES.

1. A. E. Ingham, *The Distribution of Prime Numbers* (Cambridge, 1932).
2. W. F. Osgood, *Theory of Functions of Real Variables* (Peking, 1936; New York, 1938).
3. E. C. Titchmarsh, *The Zeta-Function of Riemann* (Cambridge, 1930).
4. A. Wintner, "A note on the Riemann ξ-Function," *Journal of the London Mathematical Society*, vol. 10 (1935), pp. 82-83.
5. ———, "On a class of Fourier transforms," *American Journal of Mathematics*, vol. 58 (1936), pp. 45-90.

SMALL PERTURBATIONS.*

By AUREL WINTNER.

1. Let $A = A(t)$, where $A = (a_{ik})$ and $i, k = 1, \cdots, n$, be a matrix of n^2 functions $a_{ik} = a_{ik}(t)$ defined and continuous for $0 \leqq t < \infty$. Then the system of differential equations $x' = A(t)x$, where $x = (x_1, \cdots, x_n)$ and $x' = dx/dt$, has a unique solution $x = x(t)$, $0 \leqq t < \infty$, assigned by an arbitrary initial vector $x(0)$. Suppose that $A(t)$ becomes *practically* independent of t when t is large (the meaning of such an assumption can be made precise by a number of inequivalent definitions, two of which will be considered in a moment). Then, if A_0 denotes the constant matrix determined by that asymptotic approximation to $A(t)$, various applications assume that, in reasonable situations specific formulations of which could worry only a pure mathematician, the dissipation effected by $A(t) - A_0$ becomes negligible when $t \rightarrow \infty$.

Straightforward examples show that the exclusion of secular terms in the general solution of the approximative system $y' = A_0 y$ certainly is a *necessary* restriction. This restriction, which means the exclusion of non-linear elementary divisors for the constant matrix A_0, will always be assumed in what follows.

Certain *sufficient* conditions are known since the work of Liapounoff and Poincaré on the case of analytic differential equations (linear or not). Their results were subsequently extended by Cotton, Perron and others (for references cf. [3], p. 130). The result of Perron, generalizing Poincaré's theorem concerning the case of rational coefficients, assumes that all n^2 elements of $A(t)$ tend to finite limits as $t \rightarrow \infty$ and that, if A_0 denotes the limit matrix, the n characteristic numbers of A_0 have n distinct real parts. Under these assumptions, Perron proves that, *on a logarithmic scale,* $x' = A(t)x$ is asymptotically equivalent to $y' = A_0 y$. This does not mean asymptotic behavior proper, since $f_1(t) \sim f_2(t)$ is a weaker statement than $\log f_1(t) \sim \log f_2(t)$ as $t \rightarrow \infty$.

2. In contrast, Cotton considers instances of asymptotic behavior proper but assumes that the Poincaré-Perron requirement $A(t) \rightarrow A_0$ is replaced by

---

* Received April 11, 1945.

the requirement that the deviation of $A(t)$ from A_0 is absolutely integrable, i. e., that

$$\int^{\infty} | a_{ik}(t) - \alpha_{ik} | \, dt < \infty \tag{1}$$

holds for all n^2 elements of $A = (a_{ik})$ and $A_0 = (\alpha_{ik})$. On the other hand, due to the abandonment of the logarithmically weakened scale, only particular, rather than the general, solutions of $A(t)$ can now be discussed. The details of the proof are given for the case $n = 2$ under the assumption that one of the characteristic numbers of A_0 is positive and the other is negative ([2], pp. 479-487). However, the proof is readily seen to remain valid, in case of an arbitrary n, under the assumption that the characteristic numbers of A_0 have n distinct real parts; so that the *algebraical* assumption is the same as in Poincaré's theorem as generalized by Perron.

Except when $n = 1$, this assumption excludes every case of "stable" type (i. e., every case in which all characteristic numbers of A_0 are purely imaginary), as well as the case of every real A_0 having at least one complex characteristic number. However, Cotton explains ([2], pp. 492-493) that his method applies, if $n = 2$, to the case of a conjugate pair of purely imaginary characteristic numbers also. Actually, it is easy to see from the formal structure of the differential equations that, if $n = 2$, the case of an arbitrary conjugate pair of complex characteristic numbers is reducible to the particular case just mentioned (cf. the Remark at the end of the proof of (*) below).

3. If the n characteristic numbers have distinct real parts, the Poincaré-Perron discussion applies to n linearly independent solutions but assumes the logarithmic scale. It was observed after (1) that the latter assumption is essential, i. e., that only an m-parameter sheaf, where $1 \leqq m < n$, can in general exist in the problem of asymptotic behavior proper.

This fact can be proved by the example in which the system $x' = A(t)x$ is represented by

$$x_1' = 1 \cdot x_1 + 0 \cdot x_2 + 0, \qquad x_2' = 0 \cdot x_1 + 0 \cdot x_2 + x_1/t^2, \tag{2}$$

where $n = 2$ and $(x_1, x_2) = x$. In (2), the characteristic numbers of A_0 are real, and both $a_{ik}(t) \rightarrow \alpha_{ik}$ and (1) are satisfied. However, while the approximative system $y' = A_0 y$ belonging to (2) is $y_1' = y_1$, $y_2' = 0$ and has therefore the general solution

$$y_1(t) = c_1 e^t, \qquad y_2(t) = c_2,$$

the general solution of (2) itself is seen to be

$$x_1(t) = c_1 e^t, \qquad x_2(t) = c_2 + c_1 \int_1^t (e^u/u^2)\,du$$

$(t > 0)$. Consequently, there is no proper asymptotic relationship within the complete families of the general solutions, but only within the 1-parameter sheaves $c_1 = 0$. In fact, the complementary 1-parameter sheaves $c_2 = 0$ are such that $y_2(t)$ vanishes identically but $x_2(t)$, being an arbitrary constant multiple of the definite integral in the last formula line, tends to ∞ as $t \to \infty$ (if $c_1 \neq 0$).

Incidentally, since this integral clearly is $o(e^t)$, and therefore of a lower order than the components $y_1(t)$, $x_1(t)$, not even an involvement of the latter components could lead to an asymptotic connection between (x_1, x_2) and (y_1, y_2), except within the 1-parameter sheaf $c_1 = 0$.

Thus it is of particular interest that, if Poincaré's assumption of distinct real parts for the characteristic numbers of A_0 is replaced by its precise opposite, namely, by the assumption of "stable" characteristic numbers, a case of first interest in the applications, then (whether the purely imaginary characteristic numbers, some of which can be 0, are or are not distinct) there exists a *proper* asymptotic relation between *every* solution $x(t)$ of the system $x' = A(t)x$ and a corresponding solution $y(t)$ of the trivial system $y' = A_0 y$ (*even if* $n > 2$). This is the content of (*) below.

4. As mentioned at the very beginning, no sharp connection between the system $x' = A(t)x$ and its stationary approximation $y' = A_0 y$ can be expected if the elementary divisors of A_0 (i. e., of the constant matrix (a_{ik}) for which (1) is assumed) are allowed to be multiple. In fact, the additional assumption that all characteristic numbers of A_0 are of stable type is then of no avail. This is shown by the following example, which proves that the corresponding assumption of (*) below cannot be omitted.

Consider the scalar function

$$u(t) = \exp i(t + \log t)\,; \tag{3.}$$

so that $|u| = 1$ for every $t (> 0)$. Thus, if $u' = du/dt$,

$$u' = (i + i/t)u \quad \text{and} \quad u'' = (-1 - 2/t + \text{const.}/t^2)u.$$

Hence, if u'', u', u are multiplied by 1, α, β respectively, it follows by addition that, if α and β are so chosen as to satisfy the pair of conditions

$$-1+i\alpha+\beta=0, \qquad -2+i\alpha=0,$$

that is, if $\alpha=-2i$ and $\beta=-1$, then (3) is a solution of the scalar differential equation

$$u''-2iu'+(-1+f)u=0, \tag{4}$$

where $f=f(t)$ is $1/t^2$ times a constant. If (4) is written in the form $x'=A(t)x$, where $x=(x_1,x_2)$ and $x_1=u$, $x_2=u'$, then the assumption (1) is satisfied, since $f(t)=O(1/t^2)$, and what corresponds to $y'=A_0y$ is the scalar differential equation

$$w''-2iw'-w=0. \tag{5}$$

Consequently, if the case of multiple elementary divisors were admissible, then every solution of (4) ought to be asymptotic to an appropriate solution of (5). And this would imply the existence of two constants c_1, c_2 satisfying

$$\exp i(t+\log t) \sim (c_1+c_2t)\exp it$$

as $t\rightarrow\infty$. In fact, the function on the right of this asymptotic relation is the general solution of (5), and the function on the left is the solution (3) of (4). But the assertion of the last formula line is equivalent to

$$\exp(i\log t)\sim c_1+c_2t, \qquad t\rightarrow\infty,$$

and is therefore false for every pair of constants c_1, c_2.

5. In the particular case $x_2'=x_1$, in which the system $x'=A(t)x$ of $n=2$ linear differential equations of first order for $x=(x_1,x_2)$ is identical with one linear differential equation of second order for x_2, the content of (*) below is classical from that deduction of the asymptotic formula of Bessel's functions $J_m(t)$ which, avoiding the representation of $J_m(t)$ as a Fourier integral, is based directly on the differential equation of J_m (cf., e.g., Picard's indications in [4], pp. 417-418). However, the general case of two equations of first order cannot be reduced to one equation of second order. It is true that such a reduction is possible in a formal fashion. However, the eliminations involved assume that every coefficient function $a_{ik}(t)$ of $x'=A(t)x$, instead of being just continuous, is differentiable (and must, therefore, be required to have a continuous derivative). But even if every $a_{ik}(t)$ is restricted to be analytical, it is not clear that the fundamental assumption corresponding to (1) is satisfied for the resulting differential equation of the second order, when (1) is satisfied for the given pair of differential equations of the first order.

However, the following theorem will be proved for *every* n.

(*) *Let* $A(t) = (a_{ik})$, *where* $0 \leqq t < \infty$ *and* $i, k = 1, \cdots, n$, *be a matrix of* n^2 *continuous functions* $a_{ik} = a_{ik}(t)$ *corresponding to which there exists a matrix* $A_0 = (\alpha_{ik})$ *of* n^2 *constants* α_{ik} *satisfying* (1). *Suppose that the general solution of the differential equations* $y' = A_0 y$, *where* $y = (y_1, \cdots, y_n)$, *is a superposition of simple vibrations (i. e., that all elementary divisors belonging to* A_0 *are linear and that all characteristic numbers of* A_0, *which need not be distinct, have a common real part, which is* 0). *Then every solution* $x(t)$ *of the system* $x' = A(t)x$, *where* $x = (x_1, \cdots, x_n)$, *determines a finite almost-periodic sum* $y(t)$ *which satisfies* $x(t) - y(t) \rightarrow 0$ *as* $t \rightarrow \infty$ *and is a solution of the trivial system* $y' = A_0 y$.

Aside from the circumstance that (*) takes care of all the "stable" cases of the applications, the interest of (*) is three-fold: The asymptotic statement of (*), in contrast to that of the Poincaré-Perron theorem, refers to the *proper,* rather than to the logarithmically weakened scale and supplies, in contrast to the possibility existing in the example (2) belonging to $n = 2$, the ultimate behavior of *all* solutions of $x' = A(t)x$, although it allows *every n.*

6. The classical case of a linear differential equation of the second order which "nearly" is the equation $u'' + \lambda^2 u = 0$ of a linear oscillator represents a particular case of (*). In fact, (*) implies the following

COROLLARY. *If* $f = f(t)$, $0 \leqq t < \infty$, *is a continuous function satisfying*

$$\int^{\infty} |f(t)|\, dt < \infty, \tag{6}$$

and if λ *is a positive number, then every solution* $u(t)$ *of the differential equation*

$$u'' + (\lambda^2 + f)u = 0 \tag{7}$$

determines two constants a, b *for which*

$$u(t) - (a \cos \lambda t + b \sin \lambda t) \rightarrow 0 \tag{8}$$

holds as $t \rightarrow \infty$. *(In addition, the differentiation of this limit relation is legitimate, that is,*

$$u'(t) - (\alpha \cos \lambda t + \beta \sin \lambda t) \rightarrow 0, \tag{8 bis}$$

where $\alpha = \lambda b$, $\beta = -\lambda a$.)

In order to see this, it is sufficient to observe that, if the scalar equation (7), where $f = f(t)$, is identified with $x' = A(t)x$, and the approximation $w'' + \lambda^2 w = 0$ to (7) with $y' = A_0 y$, then the case $n = 2$ of (*) is applicable, since the assumption (1) appears in the form (6) and the constant matrix A_0 becomes

$$\begin{pmatrix} 0 & -\lambda \\ \lambda & 0 \end{pmatrix};$$

a matrix having the distinct, purely imaginary characteristic numbers $\pm \lambda i$.

It is instructive to illustrate the content of the Corollary by applying it to the differential equation,

$$v'' + v'/t + (1 - m^2/t^2)v = 0, \tag{9}$$

of Bessel's function $J_m(t)$. If $u = t^{\frac{1}{2}}v$, then (9) appears in the self-adjoint form (7), where $\lambda = 1$ and $4f = (1 - 4m^2)/t^2$. Hence, $f(t) = O(1/t^2)$ as $t \rightarrow \infty$. Since this implies that (6) is satisfied, (8) is applicable and supplies, since $\lambda = 1$ and $v = u/t^{\frac{1}{2}}$, the existence of two constants a, b for which the asymptotic formula

$$v(t) = (a \cos t + b \sin t)/t^{\frac{1}{2}} + o(1/t^{\frac{1}{2}}) \tag{10}$$

holds as $t \rightarrow \infty$. This agrees with the standard relation

$$J_m(t) = (\tfrac{1}{2}\pi t)^{-\frac{1}{2}} \cos\,(t - \tfrac{1}{2}(m + \tfrac{1}{2})\pi) + O(1/t). \tag{10 bis}$$

Actually, a comparison of (10) with (10 bis) clears up the methodical situation in two different respects:

(α) The constants a, b, for which (10 bis) supplies the values

$$a = (\tfrac{1}{2}\pi)^{-\frac{1}{2}} \cos \tfrac{1}{2}(m + \tfrac{1}{2})\pi, \qquad b = (\tfrac{1}{2}\pi)^{-\frac{1}{2}} \sin \tfrac{1}{2}(m + \tfrac{1}{2})\pi,$$

are left undetermined by the result (10). This is quite natural, since a, b in (10) are arbitrary integration constants. In fact, (10) holds not only for the particular solution $v = J_m(t)$ of (9) but for any $v = \text{const.}\, J_m(t)$ as well and, in addition, for those solutions $v(t)$ of (9) which contain the solution linearly independent of $J_m(t)$.

(β) The remainder term of (10 bis) is sharper than that of (10). This is quite natural, since, on the one hand, (10) is identical with (8) in virtue of $u = t^{\frac{1}{2}}v$ and, on the other hand, (8) is a result depending only on the general assumption (6), rather than on the drastic estimate $f(t) = O(1/t^2)$, or even on the explicit relation $f(t) = \text{Const.}/t^2$, available in the present particular case. Actually, if the assumption (6) is refined to $f(t) = O(1/t^2)$, then the $o(1/t^{\frac{1}{2}})$ in (10) can be replaced by $O(1/t)$, since the remainder term of (8), which is $o(1)$, can then be improved to $O(1/t^{\frac{1}{2}})$. The truth of this assertion, and also of its extension to the case of similar explicit assumptions (such as $f(t) = O(1/t^{1+\theta})$, where $\theta > 0$) will become clear from the proof of the general result (*), of which (8) was a corollary.

7. What concerns (*) itself, the substance of its assertion is a simple "Abelian" (in contrast to some "Tauberian") lemma. The lemma in question is that particular case of (*) which results when A_0 is chosen to be the matrix (0); a matrix which satisfies the requirements of (*), since it is a diagonal matrix in which all diagonal elements are purely imaginary, namely, 0. But if $A_0 = (0)$, then every solution $y(t)$ of $y' = A_0 y$ is independent of t. Consequently, if α_{ik}, $a_{ik}(t)$, x in (*) are replaced by 0, $f_{ik}(t)$, z respectively, there results the following fact:

(i) *If* $F(t) = (f_{ik})$, *where* $0 \leqq t < \infty$ *and* $i, k = 1, \cdots, n$, *is a matrix of* n^2 *continuous functions* $f_{ik} = f_{ik}(t)$ *satisfying*

$$\int^{\infty} |f_{ik}(t)|\, dt < \infty, \tag{11}$$

then each of the components $z_i = z_i(t)$ *of every solution* $z = z(t)$ *of the system* $z' = F(t)z$, *where* $z' = dz/dt$ *and* $z = (z_1, \cdots, z_n)$, *tends to a finite limit as* $t \to \infty$.

Actually, it will turn out that this primitive Abelian theorem contains (*) as a corollary.

It is a general principle that an Abelian theorem, which concerns a limiting equality, represents the extreme case of a less sharp but more fundamental fact, which supplies limiting inequalities (involving, if possible, absolute constants). However, (i) can hardly be improved in this direction, since, on the one hand, the assertion of (i) proves to be equivalent to the much rougher assertion

$$z(t) = O(1), \qquad (t \to \infty), \tag{12}$$

and, on the other hand, nothing less than the assumption (11) can insure that (12) holds for every solution $z(t)$ of $z' = F(t)z$ (cf. the proof of (i) below).

Consequently, all that it is reasonable to ask for is that deteriorated form of the estimate (12) which results if the assumption (11) is lightened somewhat. The simplest specific order not implying (11) is the order of $o(1/t)$ for every $f_{ik}(t)$. But all that can be said in this case is that $z(t) = O(t^{\epsilon})$ holds for every fixed $\epsilon > 0$ (rather than, as in the case (12) of (11), for $\epsilon = 0$). For the sake of possible reference in the proofs to be given below, let the positive part of this assertion be worded as a formal remark:

(i bis) *If* $F(t)$ *is a matrix of* n^2 *continuous functions which are* $o(1/t)$ *as* $t \rightarrow \infty$, *then each of the* n *components of every solution* $z(t)$ *of* $z' = F(t)z$ *is* $O(t^\epsilon)$ *for every fixed* $\epsilon > 0$.

Actually, (i bis) will turn out to be just a limiting case of (ii) below.

8. The "Abelian" nature of the above assertions suggests that some of the fundamental theorems in the classical theory of linear differential equations are disguised formulations of more general and quite primitive facts, which can be completely isolated from the theory of analytical functions of a complex variable. Such an approach can be expected to furnish not only more *general* theorems than, but also *simplified* forms for the proofs of, the classical results, since the arrangement must be so primitive as to hold under the mere assumption of continuous functions of a real variable, and to be able to retard the introduction of the specific complex-analytical restriction.

In case of Fuchs' fundamental theorem concerning "regular" singular points, the possibility of such an isolation of the underlying general facts, which have nothing to do with analyticity, is indicated by Schlesinger's proof of this theorem, as given in [5], pp. 168-170 (cf. also the presentation in [6], pp. 142-145, where reference is given to a paper of Fuhr). In this regard, it is instructive to consider Bieberbach's reproduction ([1], pp. 175-176) of Birkhoff's variant of Schlesinger's proof. In fact, the proof as presented by Bieberbach is quite incomplete, since the least attractive part of the necessary considerations, and even an observation as to the existence of such a part, is left to the ingenuity of that legendary beginner for whom text-books are written. The part in question ought to concern the differentiability of the squared absolute value of the functions involved, which are regular. This differentiability property *can*, of course, be proved. That it *needs* a proof, is clear from the fact that the absolute values themselves are not differentiable at simple zeros. *If* the latter cluster at the singular point, the proof ceases to be one. And the zeros *could* cluster, since the fact that the singular point does not correspond to an essential singularity of the solutions is just the statement to be proved and cannot therefore be used in the proof. Thus the presentation of Fuchs' theorem in [1] affords an instance of the disadvantages which can arise if facts depending on analyticity are not isolated from those having nothing to do with it.

Schlesinger's proof, as well as its subsequent variants, depends on his observation that, instead of all the lengthy formal work involved in the earlier proofs, nothing but a uniform estimate of finite order is needed for the solutions (the uniformity refers to all directions issuing from the singular point in the complex plane). But this in itself suggests the possibility of isolating a general fact of the type described above. This general fact is the content of the following theorem, the proof of which will be just as primitive as that of (i). Not even an application of Liapounoff's square-sum device (cf., e. g., [4], pp. 382-385), on the use of which Birkhoff's variant of Schlesinger's proof depends, is needed.

(ii) *If* $F(t)$ *is a matrix of* n^2 *continuous functions of the real variable* t *all of which are* $O(1/t)$ *as* $t \rightarrow \infty$, *then each of the* n *components of any solution* $z(t)$ *of the system* $z' = F(t)z$ *is* $O(t^c)$ *as* $t \rightarrow \infty$, *where* c *is a sufficiently large constant, depending only on the values of the* n^2 *constants which multiply* $1/t$ *in* $O(1/t)$.

In fact, the proof of (ii) will be such as to show that c can be chosen to be any number greater than the absolutely greatest characteristic number of the matrix (β_{ik}) defined by

$$\beta_{ik} = \limsup_{t\rightarrow\infty} | f_{ik}(t) |/t,$$

where $(f_{ik}) = F$ (according to Frobenius, this characteristic number is real and non-negative, since $\beta_{ik} \geqq 0$). In particular, if the assumption of (ii) is replaced by that of (i bis), then, since (β_{ik}) becomes the matrix (0), the constant c can be chosen to be any $\epsilon > 0$, as stated by (i bis).

Of the assertions italicized above, (*), (i), (ii) remain to be proved. It is logical to prove these three assertions in the order reverse to that in which they were stated, since their degree of coarseness is monotone decreasing when (*) is at the first place.

9. Proof of (ii). By assumption, there exists a constant matrix B satisfying the n^2 inequalities $|f_{ik}(t)| \leqq b_{ik}/t$, where $(f_{ik}) = F$, $(b_{ik}) = B$ and, without loss of generality, $1 \leqq t < \infty$. With reference to any fixed solution $z(t) = (z_1, \cdots, z_n)$ of the system $z' = F(t)z$, let $s(t) = (s_1, \cdots, s_n)$ denote that solution of the system $s' = Bs/t$ for which the n initial values $s_i(1)$ are the absolute values of the corresponding components of the initial vector $z(1)$. Let $z^m(t)$ and $s^m(t)$ denote, respectively, the infinite sequences of functions defined by the successive approximations

$$z^m(t) = z(1) + \int_1^t f(u)z^{m-1}(u)du \quad \text{and} \quad s^m(t) = s(1) + \int_1^t Bs^{m-1}(u)/u\, du,$$

where $z^1(t) = z(1)$ and $s^1(t) = s(1)$ for every $t(\geqq 1)$. Then it is clear from the definitions of B and $s_i(1)$ that $|z_i^m(t)| \leqq s_i^m(t)$ holds for every t and for every positive integer m (the subscript i refers to any of the n components of the vectors z^m, s^m).

Since both systems of differential equations are linear, so that the successive approximations $z^m(t,), s^m(t)$ converge, as $m \to \infty$, to the corresponding solutions $z(t), s(t)$ for every t, it follows from the last inequality that $|z_i(t)| \leqq s_i(t)$ holds for every t. Consequently, the assertion of (ii) will be proved if it is ascertained that each of the n components of every solution $s(t)$ of the system $s' = Bs/t$ is $O(t^c)$ as $t \to \infty$, where c is a constant depending on B.

The substitution $t = \exp t^*$ transforms the system $s' = Bs/t$, where $s' = ds/dt$, into the system $ds/dt^* = Bs$, which has constant coefficients (Euler-Cauchy). Hence, the general solution of the latter system is a linear combination of $\exp(\mu_1 t^*), \cdots, \exp(\mu_n t^*)$, if the characteristic numbers $\mu_1, \cdots, \mu_n$ of B are all distinct. If they are not distinct, there is the possibility (though not the necessity) that some of the exponentials acquire the factors $\log t^*$, $\log^2 t^*, \cdots, \log^{n-1} t^*$, each of which is $O(t^*)^\epsilon$ for every fixed $\epsilon > 0$. In any case, if μ denotes the greatest of the real parts of the n characteristic numbers μ_i, every solution s is seen to be $O \exp(ct^*)$, where $c = \mu + \epsilon$ and $\epsilon > 0$. Since $t^* = \log t$, this O-estimate completes the proof of (ii).

10. Suppose that the $O(1/t)$ of the hypothesis of (ii) is generalized to $O(1/t^\mu)$, where μ is a fixed positive index. In contrast to the "rank" in the Poincaré-Perron-Birkhoff theory of Thomé's normal series for the analytical case, μ need not be an integer. But this contrast is unessential, since the theory of "rank" can be transcribed to the so-called hypercontinuous case, in which case the coefficient functions, instead of being analytical, have derivatives of arbitrarily high order and possess formal expansions proceeding according to powers of t, which can be fractional. What *is* essential is that no such expansions need now exist for the functions occurring in the coefficient matrix, which are allowed to be arbitrary continuous functions of t.

It is clear from the proof of (ii), where $\mu = 1$, that the estimate supplied by the assertion of (ii) can in no sense be improved. On the other hand, if $\mu > 1$, then (i) is applicable, and so the estimate in the limiting case $\mu = 1$ of (ii) appears as being quite rough. However, if $\mu = 1$ is thought of as the limiting case of the case $\mu < 1$, then the estimate supplied by the assertion of (ii) becomes sharp indeed, since all that can in general be true in the case $\mu < 1$ is contained in the following fact:.

(ii bis) *If* $F(t)$ *is a matrix of* n^2 *continuous functions which, as* $t \rightarrow \infty$, *are* $O(1/t^\mu)$, *where* $0 < \mu < 1$, *then each of the n components of every solution of the system* $z' = F(t)z$ *is* $O(e^{ct^{1-\mu}})$ *as* $t \rightarrow \infty$, *where c is a sufficiently large constant, depending only on* μ *and on the* n^2 *constants multiplying* $1/t^\mu$ *in* $O(1/t^\mu)$.

In fact, the "best" majorant system for the successive approximations to the solution of $z' = F(t)z$ is now a system of the form $s' = Bs/t^\mu$, where the notations are the same as in the above proof of (ii); so that B is a real constant matrix with non-negative elements, and s is a vector $(s_1, \cdots, s_n)$. But s' denotes ds/dt, and so it is readily seen that the introduction of the (generalized) logarithmic time-scale, defined by the substitution

$$t \rightarrow (\log t)/(1-\mu), \qquad (0 < \mu < 1),$$

transforms the majorant system $s' = Bs/t^\mu$ of $z' = F(t)z$ into $s' = Cs/t$, where the last prime refers to differentiation with respect to the new t and C denotes $1/(1-\mu)$ times the constant matrix B. (The factor $1/(1-\mu)$, occurring in the definitions of the t-transformation and of C, is positive, since $0 < \mu < 1$.) But if c is any number greater than the real part of any of the n characteristic numbers of the constant matrix C, then, for that trivial reason on which the end of the above proof of (ii) depends, each of the n components of every solution $s = s(t)$ of $s' = Cs/t$ is $O(t^c)$ as $t \rightarrow \infty$.

Since the inverse of the substitution of the last formula line is the transformation

$$t \rightarrow e^{t^{1-\mu}}, \qquad (0 < \mu < 1),$$

and since the latter transforms the system $s' = Cs/t$ into the original majorant $s' = Bs/t^\mu$ of the given system $z' = F(t)z$, the estimate $O(t^c)$ of the solutions of $s' = Cs/t$ proves the exponential estimate of the assertion of (ii bis). It is also seen that this exponential estimate cannot be improved, since $F(t)$ may be exactly $1/t^\mu$ times the constant matrix B.

11. Proof of (i). Let any given solution $x(t) = (x_1, \cdots, x_n)$ of $x' = F(t)x$, where $F(t) = (f_{ik})$, be thought of as determined by the initial vector $x(c)$, where c is a fixed value of t. According to (11), it is possible to choose c so large that

$$\sum_{k=1}^{n} \int_c^\infty |f_{ik}(t)|\, dt \leqq \tfrac{1}{2} \tag{13}$$

holds for all n values of i. It will be assumed that c is fixed in this manner, and that $c \leqq t < \infty$.

Let C_m, where $m = 1, 2, \cdots$ (and, for the present, $C_m \leqq \infty$) be the least common upper bound of the n functions

$$|z_1^{m+1}(t) - z_1^m(t)|, \cdots, |z_n^{m+1}(t) - z_n^m(t)| \tag{14}$$

for $c \leqq t < \infty$, where, if $z_1(c), \cdots, z_n(c)$ are the initial constants determining a solution $z_1(t), \cdots, z_n(t)$, the functions $z_1^0(t), \cdots, z_n^0(t)$ are respectively equal to $z_1(c), \cdots, z_n(c)$ for every t and the remaining functions $z_i^m(t)$ are those assigned by the recursion formula of the successive approximations, that is, by

$$z_i^{m+1}(t) = z_i(c) + \int_c^t \sum_{k=1}^n f_{ik}(u) z_k^m(u)\,du.$$

If m is here replaced by $m-1$, it follows, by subtraction, that the absolute value of the i-th of the n functions (14) does not exceed

$$\int_c^t \sum_{k=1}^n |f_{ik}(u)|\ |z_k^m(u) - z_k^{m-1}(u)|\,du \leqq \int_c^\infty \sum_{k=1}^n |f_{ik}(u)|\, C_{m-1}\,du;$$

a bound which, according to (13), is majorized by $\frac{1}{2}C_{m-1}$. Consequently, $C_m \leqq \frac{1}{2}C_{m-1}$. Hence, $C_1 + C_2 \cdots$ is a convergent series. Since C_m majorizes all n functions (14) for every t, and since the n functions $z_i^1(t)$ are independent of t, it follows that the functions $z_1^m(t), \cdots, z_n^m(t)$, where $m = 1, 2, \cdots$ and $c \leqq t < \infty$, are uniformly bounded. This proves (12), since $z^m(t) \rightarrow z(t)$ as $m \rightarrow \infty$ holds for every t.

Although (12) claims less than (i), it implies (i) in virtue of (13). In fact, since $z(t)$ is a solution of $z' = F(t)z$,

$$z(t) = z(c) + \int_c^t F(u)z(u)\,du.$$

Hence, the assertion of (i) is equivalent to the convergence of the integral

$$\int_c^\infty F(t)z(t)\,dt.$$

But (12) and (11), where $(f_{ik}) = F$, imply the *absolute* convergence of this integral; so that the proof of (i) is complete.

12. Proof of (*). The assumptions and the assertion of (*) are invariant under transformations of x into Tx, where T denotes a constant matrix of non-vanishing determinant. Since the elementary divisors belonging to A_0 are supposed to be linear, it follows that A_0 can be assumed to be a diagonal matrix. Then, if $\lambda_1, \cdots, \lambda_n$ denote the characteristic numbers of A_0, the system $x' = A(t)x$ appears in the form

$$(14) \qquad x_i' = \lambda_i x_i + \sum_{k=1}^{n} f_{ik}(t) x_k, \qquad (i = 1, \cdots, n),$$

where $f_{ik}(t)$ represents the continuous function which, in the notations of the wording of (*), is the difference $a_{ik}(t) - \alpha_{ik}$. Hence, the assumption (1) of (*) means that (11) is satisfied by (14).

The trivial system $y' = A_0 y$ belonging to the representation (14) of $x' = A(t)x$ is the system $y_i' = \lambda_i y_i$ and has, therefore, the general solution $y_i(t) = c_i e^{\lambda_i t}$, where $i = 1, \cdots, n$. Accordingly, the assertion of (*) is equivalent to the statement that, if $(x_1, \cdots, x_n)$ is any solution of (14), each of the n functions

$$(16) \qquad z_i(t) = e^{-\lambda_i t} x_i(t)$$

tends to a finite limit, c_i, as $t \to \infty$. Consequently, it is sufficient to show that the n functions (16) represent a solution of some system $z' = F(t)z$ to which (i) is applicable. But such a system results if (14) is expressed in terms of the z_i defined by (16).

In fact, if $x_i = e^{\lambda_i t} z_i$ and its derivative x_i', which is identical with $e^{\lambda_i t}$ times $z_i' + \lambda_i z_i$, are substituted into (14), the resulting transform of (14) may be written in the form

$$(17) \qquad z_i' = \sum_{k=1}^{n} e^{(\lambda_k - \lambda_i)t} f_{ik}(t) z_k.$$

This system is of the type $z' = F(t)z$, except that the matrix element f_{ik} of $F = (f_{ik})$, instead of being the function $f_{ik}(t)$ occurring in (14), is $e^{(\lambda_k - \lambda_i)t}$ times that $f_{ik}(t)$. But the exponential function multiplying the latter $f_{ik}(t)$ is of absolute value 1, in virtue of that assumption of (*) which has not been used so far, namely, of the assumption that $\lambda_1, \cdots, \lambda_n$ are purely imaginary (possibly 0). Since (1) is satisfied by the functions $f_{ik}(t)$ occurring in (17), it follows that (1) is satisfied by the elements $f_{ik}(t)$ of the matrix $F = (f_{ik})$ also. This proves that (i) is applicable to the representation $z' = F(t)z$ of (17).

Remark. What is actually used in this proof of (*) is, not the full force of the assumption that all n characteristic numbers λ_i be purely imaginary, but merely the assumption that the n^2 exponential factors be of absolute value

1 for every t. And this will be the case whenever all n^2 differences $\lambda_i - \lambda_k$ are purely imaginary (including 0). But all that this assumption requires is an arbitrary common real part for all characteristic numbers. Thus the true content, though not the almost-periodic wording, of (*) actually is a theorem, the algebraical assumption of which is just the opposite of the corresponding assumption of Poincaré's theorem (in which the real parts of the characteristic numbers are required to be all distinct).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] L. Bieberbach, *Theorie der Differentialgleichungen*, Berlin, 1923.

[2] E. Cotton, "Sur les solutions asymptotiques des équations différentielles," *Annales Scientifiques de l'École Normale Supérieure*, ser. 3, vol. 28 (1911), pp. 473-521.

[3] O. Perron, "Über Stabilität und asymptotisches Verhalten der Integrale von Differentialgleichungssystemen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 129-160.

[4] E. Picard, *Traité d'Analyse*, 3rd ed., vol. 3, Paris, 1928.

[5] L. Schlesinger, *Vorlesungen über lineare Differentialgleichungen*, Leipzig-Berlin, 1908.

[6] ————, *Einführung in die Theorie der gewöhnlichen Differentialgleichungen auf funktionentheoretischer Grundlage*, 3rd ed., Berlin-Leipzig, 1922.

SPANS IN LEBESGUE AND UNIFORM SPACES OF TRANSLATIONS OF PEAK FUNCTIONS.*

By Ralph Palmer Agnew.

---

1. Introduction. For each $p \geqq 1$, let L_p be the Lebesgue space whose elements are the real or complex valued measurable functions $f(x)$ such that $|f(x)|^p$ is integrable over $-\infty < x < \infty$. The distance $\| f_2 - f_1 \|$ between two elements f_1 and f_2 of L_p is defined by

$$\| f_2 - f_1 \| = \{ \int_{-\infty}^{\infty} | f_2(x) - f_1(x) |^p \, dx \}^{1/p}. \tag{1.1}$$

Let E be a set in L_p. The linear manifold $M(E)$ determined by E is the set of all linear combinations (finite) of elements of E. The *span* $S_p(E)$ of E in L_p is the closure in L_p of $M(E)$.

Let $F \epsilon L_p$. For each real h, the translation $F(x+h)$ of $F(x)$ is also in L_p. Let T_F denote the set of translations of F. Wiener [4] showed that, when $F \epsilon L_1$, $S_1(T_F) = L_1$ if, and only if, the Fourier transform of F has no real zeros; and that, when $F \epsilon L_2$, $S_2(T_F) = L_2$ if, and only if, the set of zeros of the Fourier transform of F has measure 0. Segal [3] complicated matters by showing that, when $1 < p < 2$, there is an element F of L_p such that (i) the zeros of the Fourier transform of F have measure 0 and (ii) the span $S_p(T_F)$ of the translations of F does *not* include the whole space L_p. The author [1] has shown that if $p > 1$ and F is a simple step function, then $S_p(T_F) = L_p$.

It is the main object of this paper to prove the following theorem by a direct method which shows how one may approximate a given function in L_p by linear combinations of translations of a given peak function.

Theorem 1.2. *If $p > 1$ and F is a peak function, then $S_p(T_F) = L_p$.*

Uniform approximation and truncated peak functions are treated in **6** and **7**.

2. Peak functions. In terms of positive constants a and b, a *peak function* $F(x)$ of *width* $2b$ and *height* ab is defined by the formula

---

* Received March 14, 1945; Presented to the American Mathematical Society, April 27, 1945.

$$(2.1) \qquad \begin{aligned} F(x) &= a(b - |x|) \qquad && |x| \leqq b \\ &= 0 && |x| > b. \end{aligned}$$

The Fourier transform $G(u)$ of this function $F(x)$ is the continuous non-negative function for which $G(0) = ab^2(2\pi)^{-1/2}$ and

$$(2.2) \qquad G(u) = (1/2\pi)^{1/2} \int_{-\infty}^{\infty} F(x) e^{iux} dx = (2/\pi)^{1/2} a \frac{1 - \cos bu}{u^2}$$

when $u \neq 0$. Wiener's results imply that $S_1(T_F) \neq L_1$ and $S_2(T_F) = L_2$.

3. A lemma. In our proof of Theorem 1. 2, we shall use the following lemma three times.

LEMMA 3. 1. *Let $p > 1$. Let $f(x)$ be a bounded measurable function vanishing outside the interval $c - d < x < c + d$, and let $f(x + d) = f(x)$ [or $f(x + d) = -f(x)$] when $c - d \leqq x \leqq c$. Let $g(x) = f(x)$ when $c - d \leqq x \leqq c$ and $g(x) = 0$ otherwise. Then $S_p(T_f)$ contains g.*

Let $\epsilon > 0$. We establish the lemma by giving an explicit method for selecting a linear combination f_ϵ of translations of f such that $\| g - f_\epsilon \| < \epsilon$. Let M be a constant such that $|f(x)| \leqq M$ for each x. Let n be the least positive integer such that

$$(3.2) \qquad Md^{1/p}(n + 1)^{1/p} n^{-1} < \epsilon.$$

Let

$$(3.3) \qquad f_\epsilon(x) = f(x) + \sum_{k=1}^{n} (-1)^{k+\delta}(1 - k/n) f(x - kd)$$

where $\delta = 0$ if $f(x + d) = f(x)$ and $\delta = 1$ if $f(x + d) = -f(x)$. Then $|g(x) - f_\epsilon(x)|$ is zero except over $n + 1$ intervals, each of length d, where it is at most M/n. Therefore

$$(3.4) \qquad \| g - f_\epsilon \| \leqq \{ \int_0^{(n+1)d} (M/n)^p dx \}^{1/p} = Md^{1/p}(n + 1)^{1/p} n^{-1} < \epsilon$$

and the lemma is proved.

4. Approximation to narrower peak functions. We shall show how to approximate narrower peak functions by linear combinations of translations of a given one by proving the following lemma.

LEMMA 4. 1. *Let $p > 1$. Let $F(x)$ be the peak function of width $2b$ defined by (2. 1), and let $F_1(x)$ be the narrower peak function, of width $b/2$, defined by*

$$F_1(x) = a(b/4 - |x|) \qquad |x| \leqq b/4 \tag{4. 2}$$
$$= 0 \qquad |x| > b/4.$$

Then $S_p(T_F)$ contains F_1.

Let $b_1 = b/4$ and let

$$f_1(x) = F(x + 2b_1) + F(x - 2b_1) - F(x). \tag{4. 3}$$

Let $g_1(x)$ be the function defined by

$$\begin{aligned} g_1(x) &= a(x + 6b_1) && -6b_1 \leqq x \leqq -4b_1 \\ &= 2ab_1 && -4b_1 \leqq x \leqq -2b_1 \\ &= -ax && -2b_1 \leqq x \leqq 0 \\ &= 0 && \text{otherwise.} \end{aligned} \tag{4. 4}$$

The hypotheses of Lemma 3. 1 are satisfied when $f = f_1$, $g = g_1$, $c = 0$, and $d = 6b_1$. Hence the span of the translations of f_1 includes g_1 and therefore the span of the translations of F includes g_1.

Let

$$f_2(x) = F(x) - 2g_1(x - 3b_1) \tag{4. 5}$$

and

$$\begin{aligned} g_2(x) &= a(x + 4b_1) && -4b_1 \leqq x \leqq -3b_1 \\ &= -a(x + 2b_1) && -3b_1 \leqq x \leqq -b_1 \\ &= ax && -b_1 \leqq x \leqq 0 \\ &= 0 && \text{otherwise.} \end{aligned} \tag{4. 6}$$

The hypotheses of Lemma 3. 1 are satisfied when $f = f_2$, $g = g_2$, $c = 0$ and $d = 4b_1$. Hence the span of the translations of f_2 includes g_2 and therefore the span of the translations of F includes g_2.

Finally, the hypotheses of Lemma 3. 1 are satisfied when $f(x) = g_2(x)$

and $g(x) = F_1(x + 3b_1)$. Hence the span of the translations of g_2 includes $F_1(x + 3b_1)$; therefore the span of the translations of F includes $F_1(x + 3b_1)$ and hence also $F_1(x)$. This completes the proof of Lemma 4. 1.

By repeated application of Lemma 4. 1, and use of the fact that if $S_p(T_f)$ contains g then $S_p(T_f)$ contains $S_p(T_g)$, we obtain the following lemma.

LEMMA 4. 7. *If F is a peak function of width $2b$ and G is a peak function of width*

$$\delta = 2b/4^m \tag{4. 8}$$

where m is a positive integer, then $S_p(T_F)$ contains G.

5. Proof of Theorem 1. 2. Let $p > 1$, let $f \in L_p$, and let F be a given peak function defined by (2. 1). Let $\epsilon > 0$. By a fundamental theorem on approximation, there is a continuous function $f_1(x)$, vanishing outside some finite interval $-\alpha \leqq x \leqq \alpha$, such that $\| f - f_1 \| < \epsilon/3$. We may choose a positive number δ_1 such that the number δ defined by $\delta = 2\delta_1$ has the form (4. 8), and moreover the number δ_1 is so small that $\| f_2 - f_1 \| < \epsilon/3$ when $f_2(x)$ is the polygonal function for which $f_2(n\delta_1) = f_1(n\delta_1)$ when $n = 0, \pm 1, \pm 2, \cdots$ and which is linear over the interval $n\delta_1 \leqq x \leqq (n+1)\delta_1$ for each $n = 0, \pm 1, \pm 2, \cdots$. The polygonal function is a finite linear combination of peak functions of width $\delta = 2\delta_1$; in fact if

$$\begin{aligned} G(x) &= \delta_1 - |x| \qquad & |x| \leqq \delta_1 \\ &= 0 & |x| > \delta_1, \end{aligned} \tag{5.1}$$

then

$$f_2(x) = \sum f_2(n\delta_1) G(x - n\delta_1), \tag{5. 2}$$

the sum ranging over the finite set of values of n for which $f_2(n\delta_1) \neq 0$. Since Lemma 4. 7 implies that $S_p(T_F)$ contains $G(x)$, it follows from (5. 2) that $S_p(T_F)$ contains $f_2(x)$. Hence there is a linear combination $F_\epsilon(x)$ of translations of $F(x)$ such that $\| f_2 - F_\epsilon \| < \epsilon/3$. Our inequalities imply that $\| f - F_\epsilon \| < \epsilon$. Hence $S_p(T_F)$ contains f and Theorem 1. 2 is proved.

6. Uniform approximation. A modification of the above proof of Theorem 1. 2 furnishes a proof of the following theorem on uniform approximation.

THEOREM 6.1. *Let $F(x)$ be a peak function defined by* (2.1). *Let $f(x)$ be a continuous function for which $f(x) \to 0$ as $x \to \infty$ and as $x \to -\infty$. Then to each $\epsilon > 0$ corresponds a linear combination $F_\epsilon(x)$ of translations of $F(x)$ such that*

$$|f(x) - F_\epsilon(x)| < \epsilon \qquad -\infty < x < \infty.$$

The author is indebted to the referee who pointed out to him that this theorem is implied by results of Boas and Bochner [2]. In the latter paper, the italicized statement on page 287 deals with functions $f(t)$, $g(t)$, and a weight function $p(t)$. If we assume that $p(t) \equiv 1$ and that $g(t) \epsilon L$, then g has an ordinary Fourier transform and the approximation

$$|f(t) - \sum_{j=1}^{k} A_j g(t + \lambda_j)| < \epsilon$$

is possible whenever (i) $f(t)$ and $g(t)$ are continuous, except for discontinuities of the first kind in $g(t)$, and have limits 0 for $x \to \pm \infty$ and (ii) there is no interval over which the Fourier transform of $g(t)$ vanishes everywhere.

7. Truncated peak functions. Lemma 3.1 and Theorem 1.2 provide a basis for a short proof of the following theorem.

THEOREM 7.1. *Let $p > 1$. Let $\phi(x)$ be the truncated peak function defined in terms of positive constants a, b, and q for which $q < b$ by the formula*

$$(7.2) \qquad \begin{aligned} \phi(x) &= a(b - q) & |x| &\leqq q \\ &= a(b - |x|) & q \leqq |x| &\leqq b \\ &= 0 & |x| &\geqq b. \end{aligned}$$

Then $S_p(T_\phi) = L_p$.

To prove this, let

$$(7.3) \qquad \phi_1(x) = \phi(x) + \phi(x + b + q)$$

and

$$(7.4) \qquad f(x) = \phi_1(x) - \phi_1(x - b + q).$$

Then, when $g(x)$ is a translation of the peak function (2.1), the hypotheses of Lemma 3.1 hold. Hence, by Lemma 3.1, $S_p(T_f)$ includes g and therefore

$S_p(T_\phi)$ includes g. Since, by Theorem 1.2, $S_p(T_g) = L_p$ it follows that $S_p(T_\phi) = L_p$ and Theorem 7.1 is proved.

CORNELL UNIVERSITY,
ITHACA, N. Y.

BIBLIOGRAPHY

1. Agnew, R. P. "Spans in Lebesgue and uniform spaces of translations of step functions," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 229-233.
2. Boas, R. P. and Bochner, S. "Closure theorems for translations," *Annals of Mathematics* (2), vol. 39 (1938), pp. 287-300.
3. Segal, I. E. "The span of the translations of a function in a Lebesgue space," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 165-169.
4. Wiener, N. "Tauberian theorems," *Annals of Mathematics* (2), vol. 33 (1932), pp. 1-100.

SOLUTION OF A PROBLEM OF G. KOETHE.*

By Jakob Levitzki.

1. Preliminaries. The following problem has been proposed by G. Koethe:[1] Is it possible to deduce the nilpotency of a nil-ring by assuming the maximal condition (in short: M-condition)[2] for the r. ideals? This problem has been generalized by R. Baer, who says [1, p. 553] "it seems to be an open question whether or not the maximum condition for r. ideals is sufficient for nilpotence of the upper radical." In the present note[3] we answer the question of Baer, and by implication also that of Koethe, in the affirmative. Moreover, since it is not known whether, in general, the upper radical contains all one-sided nil-ideals of the ring, it is of interest to note that the M-condition for the r. ideals implies the nilpotency of all one-sided nil-ideals of the ring (compare Theorem 5 of the present note) and thus ensures the existence and the nilpotency of the radical in the sense of Koethe [3, p. 169]. An analogous result concerning the minimal condition (in short: m-condition) has been proved by C. Hopkins [2, § 3].

Turning to the more general problem concerning the nilpotency of subrings, first note that the following useful lemma can be easily verified:

Lemma. *If A is a nilpotent ideal of a ring S, and if each nil-subring of S/A is nilpotent, then also each nil-subring of S is nilpotent.*

Applying this lemma to a semi-primary ring (i. e., a ring S with a nilpotent radical N and with a semi-simple quotient-ring S/N), it follows directly from the author's theorem [4, § 2] that each nil-subring of a semi-primary ring is nilpotent (see [2, § 3]). On the other hand, applying our lemma to the final result of the present note (Theorem 6) we obtain the following theorem: *If A is a nilpotent ideal of a ring S, and if the M-condition holds for the r. ideals as well as for the l. ideals of S/A, then each nil-subring of S is nilpotent.* This last result is a considerable extension of

---

* Received January 18, 1945.

[1] G. Koethe [3, p. 165]. The numbers in brackets refer to the bibliography at the end of the paper.

[2] The abbreviations used in this paper are listed at the end of 1.

[3] Some of the results of the present note were proved in the author's paper "On multiplicative systems" which was actually in the press for publication in the *Compositio Mathematica* in 1939; it does not seem to have appeared, presumably owing to war conditions.

the aforementioned result concerning semi-primary rings since, as is well known, the M-condition holds in a semi-simple ring for both r. ideals and l. ideals.[4]

Since an extensive use will be made here of the author's former results [5, § 2], it is convenient to adopt the terminology introduced in that paper. Thus a ring T is termed *semi-nilpotent,* if each subring of T which is generated by a finite set of elements, is nilpotent. A ring which is not semi-nilpotent is called *semi-regular.* The sum N of all semi-nilpotent ideals of a ring S is defined as the *radical* of the ring. The radical N is a semi-nilpotent ideal; it is a radical-ideal in the sense of Baer [1, § 1]; it contains all one-sided semi-nilpotent ideals of the ring, and the radical of S/N is zero. Finally, if $a \in S$, then $a \in N$ if and only if $aS \subseteq N$ (or $Sa \subseteq N$).

We shall use the notion of the right quotient $(A:K)_r$ which is defined, for an ideal A and an arbitrary subset K of S, as the set of all elements x of S satisfying the relation $K \cdot x \subseteq A$. This set is evidently a r. ideal in S. Similarly the left quotient $(A:K)_l$ is defined as the set of all elements x satisfying the relation $x \cdot K \subseteq A$. An immediate consequence of the definition is the following property of the quotient: If $K \subseteq L$, then $(A:K)_r \supseteq (A:L)_r$, and $(A:K)_l \supseteq (A:L)_l$.

The following abbreviations are used in the present note: Maximal condition = M-condition; minimal condition = m-condition; two-sided ideal = ideal; right ideal = r. ideal; left ideal = l. ideal.

2. Semi-nilpotency of nil-rings. The theorems of this section are based on the properties of the radical [5, § 2] which were described in the previous section.

THEOREM 1. *Let S be an arbitrary ring, T a semi-regular nil-subring of S and N the radical of S. Then there exists an element a in T such that, on putting $R = a \cdot T$ and $L = T \cdot a$, we have:*

1) *The rings $R, L, R \cdot S$ and $S \cdot L$ are again semi-regular.*

2) $T \cdot S \supset R \cdot S,\ S \cdot T \supset S \cdot L.$

3) $(N:T)_l \subset (N:R)_l,\ (N:T)_r \subset (N:L)_r.$

4) *The l. ideal $(N:R)_l$ and the r. ideal $(N:L)_r$ are semi-regular.*

Proof. Denote by N^* the radical of T. Since T is semi-regular and N^* is semi-nilpotent, we have $N^* \subset T$. Now choose an element a such that $a \in T$,

---

[4] Further extensions of this result and others related to it will be discussed in a subsequent paper.

$a \notin N^*$; then it follows that a has the required properties. Indeed, on putting $R = a\cdot T$, $L = T\cdot a$, it follows (see [5, Theorem 5]) that R, as well as L, is semi-regular. Since the radical of T/N^* is zero, R^2 and L^2 must also be semi-regular, which in view of $R^2 \subseteq R\cdot S$, $L^2 \subseteq S\cdot L$ implies that $R\cdot S$ and $S\cdot L$ are also semi-regular. To prove 2) note that $0 \neq T\cdot S \supseteq R\cdot S$ and that from $T\cdot S = R\cdot S = a\cdot T\cdot S$ it would follow by left multiplication that $T\cdot S = a^k\cdot T\cdot S \neq 0$ for each k, which would imply that $a^k \neq 0$ for each k, in contradiction to the assumption that T was a nil-subring. Hence $T\cdot S \supset R\cdot S$, and similarly $S\cdot T \supset L\cdot T$. Turning to the proof of 3) first note that from $T \supset R$ it follows that

$$(N:T)_l \subseteq (N:R)_l. \tag{1}$$

Since, further, a is nilpotent, and $a\cdot T \not\subseteq N$ it follows that an integer k exists such that $k \supset 1$, $a^kT \subseteq N$ and $a^{k-1}\cdot T \not\subseteq N$, which implies that

$$a^{k-1} \notin (N:T)_l, \qquad a^{k-1} \in (N:R)_l. \tag{2}$$

From (1) and (2) follows the first part of 3). The second part of 3) can be proved similarly. As to the proof of 4), it follows from the definition of the quotient that $N \subseteq (N:T)_l$, and hence from the first inequality under 3) that we have $N \subset (N:R)_l$, which implies that $(N:R)_l$ is semi-regular. The semi-regularity of $(N:L)_r$ follows similarly.

THEOREM 2. *If in a ring S the M-condition or the m-condition holds for the semi-regular r. ideals then each nil-subring T of S is semi-nilpotent.*

Proof. Denote by N the radical of S, and suppose that T is a semi-regular ring, then by Theorem 1 there exist semi-regular subrings R_1 and L_1 in T such that $R_1\cdot S$ and $(N:L_1)_r$ are semi-regular r. ideals in S satisfying the relations $T\cdot S \supset R_1S$, $(N:T)_r \subset (N:L_1)_r$. Now by replacing in Theorem 1 the ring T by R_1 (or by L_1), we similarly deduce the existence of semi-regular subrings R_2 and L_2 such that $R_1S \supset R_2S$ and $(N:L_1)_r \subset (N:L_2)_r$. Proceeding in this manner we obtain, by induction, the infinite descending chain of semi-regular r. ideals

$$R_1S \supseteq R_2S \supseteq \cdots \tag{3}$$

and the infinite ascending chain of semi-regular r. ideals

$$(N:L_1)_r \subset (N:L_2)_r \subset \cdots. \tag{4}$$

By (3) and (4) it follows that as well the m-condition as the M-condition is incompatible with the supposed semi-regularity of T, i. e., that T is semi-nilpotent.

Since, in particular, either the M-condition or the m-condition of Theorem 2 implies that all nil-r. ideals and all nil-l. ideals are semi-nilpotent, it follows from Koethe's definition of the radical [3, § 3] and from Baer's definition of the upper radical [1, § 1] that

THEOREM 3. *If in a ring S the M-condition or the m-condition holds for the semi-regular r. ideals, then the radical in the sense of Koethe exists and is equal to the radical N of the ring. The radical N is then also identical with the upper radical in the sense of Baer.*

3. Chain-conditions for nilpotency of nil-subrings.

In this section we shall prove *inter alia* the statements made in Section 1.

THEOREM 4. *If in a ring S the M-condition holds for the semi-nilpotent r. ideals, then the radical N of the ring S is nilpotent.*

Proof. From the M-condition it follows that N contains a finite set of elements $a_1, a_2, \cdots, a_m$ such that $N^2 = (a_1N, a_2N, \cdots, a_mN)$. Now denote by A the ring generated by the a_i; then we evidently have $N^2 = A \cdot N$. Multiplying this equation by N we obtain $N^3 = AN^2 = A^2N$. Proceeding in this manner with right multiplication by N we obtain, by induction,

$$N^k = A^{k-1}N, \qquad (k = 2, 3, \cdots). \tag{5}$$

Now since N is a semi-nilpotent ring, and since A is generated by a finite set of elements belonging to N, it follows that A is nilpotent, and by (5) that N also is nilpotent.

Remark. In a previous note it has been shown by the author [6, Theorem 4] that the nilpotency of N is also a consequence of the m-condition for the two-sided semi-nilpotent ideals of the ring. The foregoing theorem is an extension of the well-known fact that from the maximum condition for the r. ideals it follows that the sum of all nilpotent r. ideals is itself nilpotent.

The following theorem includes the solution of the problems of Baer and Koethe which were discussed in Section 1.

THEOREM 5. *Let N denote the radical of a ring S. If the M-condition holds for the r. ideals of S containing N (i. e., for the r. ideals of S/N) and for the r. ideals contained in N (i. e., for the semi-nilpotent r. ideals of S), then all nil-r. ideals and all nil-l. ideals of S are nilpotent.*

Proof. Denote by T any nil-r. ideal or nil-l. ideal of S; then $(T, N)/N$ is a nil-r. ideal or a nil-l. ideal in S/N. Now the radical of S/N is zero [5, Theorem 3], i. e., each non-zero one-sided ideal of S/N is semi-regular, and since by Theorem 2 the nil-subring $(T, N)/N$ of S/N must be semi-nilpotent, it follows that $(T, N)/N$ must be the zero ideal of S/N, i.e., $T \subseteq N$. Now it follows, in view of Theorem 4, from the second condition of our theorem, that N is nilpotent, and since T is a subset of N, that T also is nilpotent.

Remark. An analogous theorem for the m-condition can be derived as a consequence of the author's former result [6, Theorem 5]. However, we can now improve this result as follows:

THEOREM 5a. *Let N denote the radical of a ring S. If the m-condition holds for the nil-r. ideals containing N, and for the ideals contained in N, then all nil-r. ideals and all nil-l. ideals of S are nilpotent.*

Proof. Indeed it is easily verified that if the semi-regular subring T of Theorem 1 is a nil-r. ideal or a nil-l. ideal, then the rings $R \cdot S$ and $S \cdot L$ of that theorem are a nil-r. ideal and a nil-l. ideal, respectively. Accordingly, we may assume that in this case the r. ideals of the descending chain (3) in the proof of Theorem 2, are all nil-r. ideals. These remarks, combined with the author's former result [6, Theorem 4] yield, as in the proof of Theorem 5, the required result.

Turning now to the problem of nilpotency of subrings, we prove

THEOREM 6. *If the M-condition holds for the r. ideals as well as for the l. ideals of a ring S, then each nil-subring T of S is nilpotent.*

Proof. It follows from the M-condition for r. ideals that there exists in T a finite set of elements $a_1^{(1)}, a_2^{(1)}, \cdots, a_{n_1}^{(1)}$ such that $T \cdot S = (a_1^{(1)}S, a_2^{(1)}S, \cdots, a_{n_1}^{(1)}S)$. Denote by A_1 the ring generated by the $a^{(1)}$; then we evidently have $T \cdot S = A_1 S$. Similarly it follows that T contains a finite set of elements $a_1^{(2)}, a_2^{(2)}, \cdots, a_{n_2}^{(2)}$ such that $T^2S = (a_1^{(2)}TS, \cdots, a_{n_2}^{(2)}TS)$. Denoting by A_2 the ring generated by the $a^{(2)}$, we obtain $T^2S = A_2T \cdot S = A_2 \cdot A_1 \cdot S$. Proceeding in this manner we obtain, by induction, an infinite sequence of rings $A_1, A_2, \cdots$, each of which is generated by a finite set of elements of T, such that we have

$$T^kS = A_kA_{k-1} \cdots A_1S, \qquad (k = 1, 2, \cdots). \tag{6}$$

Now suppose that

$$T^k \neq 0, \qquad (k = 1, 2, \cdots); \tag{7}$$

then by (6) in view of $T^kS \supseteq T^{k+1}$ we would have

$$A_kA_{k-1} \cdots A_1 \neq 0, \qquad (k = 1, 2, \cdots). \tag{8}$$

Since, by Theorem 2, the ring T is semi-nilpotent, it follows by the definition of the A that each A_k is nilpotent. Hence by (8) we deduce the existence of a positive integer $s_k = s$ such that

$$A_k^s A_{k-1} \cdots A_1 S = 0, \quad A_k^{s-1} A_{k-1} \cdots A_1 S \neq 0, \quad s > 1,$$

which implies that

$$(9) \qquad A_k^{s-1} \not\subseteq (0: A_{k-1} \cdots A_1 S)_l, \qquad A_k^{s-1} \subseteq (0: A_k A_{k-1} \cdots A_1 S)_l,$$

i. e., $(0: A_{k-1} \cdots A_1 S) \neq (0: A_k A_{k-1} \cdots A_1 S)$. On the other hand, it follows from (6) in view of $T^k \cdot S \subseteq T^{k-1} \cdot S$ that

$$(10) \qquad (0: A_{k-1} \cdots A_1 S)_l \subseteq (0: A_k \cdots A_1 S)_l.$$

By (9) and (10 we finally obtain

$$(11) \qquad (0: A_{k-1} \cdots A_1 \cdot S)_l \subset (0: A_k A_{k-1} \cdots A_1 \cdot S)_l, \qquad (k = 2, 3, \cdots),$$

i. e., an infinite ascending chain of l. ideals, which is a contradiction to the M-condition for l. ideals. Hence, the supposition (7) is false, i. e., T is nilpotent.

Remark. Applying the lemma of **1** we can extend Theorem 6 if we replace its conditions by assumptions analogous to those of Theorem 5.

HEBREW UNIVERSITY, JERUSALEM.

BIBLIOGRAPHY

1. R. Baer, "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.
2. C. Hopkins, "Nil-rings with minimal condition for admissible left ideals," *Duke Mathematical Journal*, vol. 4 (1938), pp. 664-667.
3. G. Koethe, "Die Struktur der Ringe deren Restklassenring nach dem Radical vollstaendig reduzibel ist," *Mathematische Zeitschrift*, Bd. 32 (1930), pp. 161-186.
4. J. Levitzki, "Ueber nilpotente Unterringe," *Mathematische Annalen*, Bd. 105 (1931), pp. 620-627.
5. ———, "On the radical of a general ring," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 462-466.
6. ———, "Semi-nilpotent ideals," *Duke Mathematical Journal*, vol. 10 (1943), pp. 553-556.

UNIVERSAL FUNCTIONS OF POLYGONAL NUMBERS, III.*

By L. W. Griffiths.

1. Introduction. The universal functions of polygonal numbers of order $m+2$, with the sum of the coefficients in each function equal to $m+4$, are determined in this paper. If the sum of the coefficients is at most $m+3$ the universal functions have been determined.[1] A necessary and sufficient condition that the universality of a function of weight $m+4$ be not implied by that of a function of weight at most $m+3$ is proved. For each m such that $m \geqq 3$ a universal function of weight $m+4$ satisfying this condition is exhibited.

In this paper use is made of facts on simultaneous representation due to Dickson,[2] and new results are proved for the case $(1, 2, 2, 3)$ with b divisible by four.

The general development of this paper is suggested by the first paper to which reference has been made. Certain notations and facts established in that paper are used. Many proofs are so similar that they are not presented here.

The polygonal numbers of order $m+2$ are defined, for m a positive integer, by $p(x) = x + m(x^2 - x)/2$ with $x = 0, 1, 2, \cdots$. Here $m \geqq 3$, for the reasons stated in I. The coefficients $a_1, \cdots, a_n$ in the functions $f = (a_1, \cdots, a_n) = a_1p_1 + \cdots + a_np_n$ are positive integers to be determined. Also $1 \leqq a_1 \leqq \cdots \leqq a_n$, and, by definition, $w_k = a_1 + \cdots + a_k$ $(1 \leqq k \leqq n)$ and $w = w_n$. It will be proved that, if $w = m + 4 \geqq 7$, then f is universal only if f satisfies (1), or (2), or (3), or (4), or (5) and (7), or (6) and (7). It will also be proved that f is universal if f satisfies (1) or (2) or (4). If f satisfies (3), or (5) and (7), or (6) and (7), then f represents every positive integer A except perhaps when A is in the specific range stated in Theorem 6. This range depends only on m and f. The methods of this paper are applicable to the verification that f actually represents these integers A. Verification is practically certain, and is being carried forward as assistance is available.

---

* Received November 30, 1944.

[1] L. W. Griffiths, *Annals of Mathematics*, vol. 31 (1930), pp. 1-12; *American Journal of Mathematics*, vol. 66 (1944), pp. 97-100. These papers will be cited as I and II respectively.

[2] L. E. Dickson, *American Journal of Mathematics*, vol. 56 (1934), pp. 513-528.

2. Necessary and sufficient conditions that f represent integers less than $145m+25$. The following lemmas and theorems have been proved by methods similar to those used in the proofs of the corresponding facts in I.

LEMMA 1. *If* $w=m+4$, *then* $f=0,\cdots,m+1$ *if and only if* $f=(1,a_2,\cdots,a_n)$ *and* $a_k\leqq w_{k-1}+1$ $(2\leqq k\leqq n)$.

LEMMA 2. *If* $w=m+4$ *and if* $f=(1,1,\cdots,a_n)$, *then* $f=m+5,\cdots,2m+3$ *if and only if* $a_k\leqq w_{k-1}$ $(3\leqq k\leqq n)$.

LEMMA 3. *If* $w=m+4\geqq 8$ *and if* $f=(1,1,\cdots,a_n)$, *then* $f=2m+6,\cdots,3m+2$ *if and only if* $a_3=1$ *or* 2 *and* $a_k\leqq w_{k-1}-1$ $(4\leqq k\leqq n)$.

Theorem 1 follows from Lemmas 1, 2, and 3, and direct verification for the integers $m+2$, $m+3$, $m+4$, $2m+4$, $2m+5$, $3m+3$, if $m\geqq 4$; if $m=3$ it follows from Lemmas 1 and 2, and direct verification for the integers $2(m+2),\cdots,3m+3$.

THEOREM 1. *If* $w=m+4\geqq 7$ *and if* $f=(1,1,\cdots,a_n)$, *then* $f=0,\cdots,3m+3$ *if and only if* $a_3=1$ *or* 2 *and* $a_k\leqq w_{k-1}-1$ $(4\leqq k\leqq n)$.

If f satisfies the conditions of Theorem 1 with $a_3=1$, then it represents the integers $3m+4,\cdots,6m+8$ if and only if it is not one of the functions $(1,1,1,1,3,\cdots,a_n)$ with $n\geqq 6$. Each of these functions does not represent $5m+9$. If f satisfies the conditions of Theorem 1 with $a_3=2$, then $f=3m+4,\cdots,5m+7$ if and only if f is the function $(1,1,2,3)$ or one of the functions $(1,1,2,a_4,\cdots,a_n)$ with $a_4=2$ or 3 and $a_k\leqq w_{k-1}-2$ $(5\leqq k\leqq n)$. If f satisfies the conditions of Theorem 1 with $a_3=2$ and if there is an integer k such that $k\geqq 5$ and $a_k=w_{k-1}-1$, then f does not represent $4m+6+a_k+\cdots+a_n$. Each of the functions $(1,1,2,2,a_5,\cdots,a_n)$ with $a_k\leqq w_{k-1}-2$ $(5\leqq k\leqq n)$ represents $5m+8,\cdots,6m+8$. Each of the functions $(1,1,2,3,a_5,\cdots,a_n)$ with $a_k\leqq w_{k-1}-2$ $(5\leqq k\leqq n)$ does not represent $5m+9$. The function $(1,1,2,3)$ has $w=7$, and hence $m=3$ and $5m+7=6m+4$; it represents $6m+5,\cdots,6m+8$ by direct verification.

THEOREM 2. *If* $w=m+4\geqq 7$ *and if* $f=(1,1,\cdots,a_n)$, *then* $f=0,\cdots,6m+8$ *if and only if* f *satisfies* (1), (2), (3), *or* (4):

(1) $f=(1,1,1,1,a_5,\cdots,a_n)$, $a_k\leqq w_{k-1}-1$ $(5\leqq k\leqq n)$, *and* $a_5\neq 3$ *if* $n>5$,

(2) $f=(1,1,1,2,a_5,\cdots,a_n)$, $a_k\leqq w_{k-1}-1$ $(5\leqq k\leqq n)$,

(3) $f=(1,1,2,2,a_5,\cdots,a_n)$, $\quad a_k\leqq w_{k-1}-2\ (5\leqq k\leqq n)$;

(4) $f=(1,1,2,3)$.

LEMMA 4. *If* $w=m+4$ *and if* $f=(1,2,a_3,\cdots,a_n)$, *then* $f=3m+4$ *if, and only if,* $a_3=2$.

LEMMA 5. *If* $w=m+4$ *and if* $f=(1,2,a_3,\cdots,a_n)$, *then* $f=m+5,\cdots,2m+3$ *if and only if there is an integer* J *such that* $2\leqq J\leqq n-1$ *and* $a_{J+1}=3$, *and that if* $J<n-1$ *then* $a_k\leqq w_{k-1}-2$ $(J+2\leqq k\leqq n)$, *and that if there is an integer* k *such that* $k>J+1$ *and* $a_k=w_{k-1}-3$ *then there is an integer* K *such that* $a_{K+1}=a_k+1$.

If f satisfies the conditions of Lemmas 4 and 5 then, by the hypothesis that the coefficients are arranged so that $a_1\leqq a_2\leqq\cdots\leqq a_n$, it is true that f satisfies the conditions of Lemma 1. By direct verification, such an f represents $m+2$, $m+3$, $m+4$, $2m+4$, $2m+5$, $3m+3$.

LEMMA 6. *If* $w=m+4$ *and if* $f=(1,2,a_3,\cdots,a_n)$, *then* $f=2m+6,\cdots,3m+2$ *if and only if* $a_k\leqq w_{k-1}-1\ (3\leqq k\leqq n)$.

Theorem 3 follows from Lemmas 1, 4, 5, and 6.

THEOREM 3. *If* $w=m+4\geqq 7$ *and if* $f=(1,2,a_3,\cdots,a_n)$, *then* $f=0,\cdots,3m+4$ *if, and only if,* $a_3=2$ *and there is an integer* J *such that* $3\leqq J\leqq n-1$ *and* $2=a_3=\cdots=a_J$ *and* $a_{J+1}=3$ *and that the remaining coefficients satisfy the conditions of Lemma* 5.

If f satisfies the conditions of Theorem 3 and if there is an integer k such that $k\geqq J+2$ and $a_k=w_{k-1}-2$, then f does not represent $4m+7+a_k$. If f satisfies the conditions of Theorem 3 with $a_k\leqq w_{k-1}-3\ (J+2\leqq k\leqq n)$, then $f=3m+5,\cdots,7m+10$. If the function $(1,2,2,3,a_5,\cdots,a_n)$ satisfies the conditions of Theorem 3 and has $a_k\leqq w_{k-1}-3\ (J+2\leqq k\leqq n)$, then it represents $7m+12$ if, and only if, there is an integer S such that $S\geqq 5$ and $a_S=4$. The function $(1,2,2,3)$ represents $7m+12$ but does not represent $8m+12$. If the function $(1,2,2,2,a_5,\cdots,a_n)$ satisfies the conditions of Theorems 3 and has $a_k\leqq w_{k-1}-3\ (J+2\leqq k\leqq n)$, then it represents the integers $7m+11,\cdots,17m+14$. If the function $(1,2,2,3,a_5,\cdots,a_n)$ satisfies the conditions of Theorem 3 and if there is an integer S such that $S>4$ and $a_S=4$, and either $S=n$ or $a_k\leqq w_{k-1}-3\ (S+1\leqq k\leqq n)$, then it represents the integers $7m+11,\cdots,17m+14$. If the function $(1,2,2,2,3,7,a_7,\cdots,a_n)$ with $n\geqq 7$ satisfies the conditions of Theorem 3 and has $a_k\leqq w_{k-1}-3\ (J+2\leqq k\leqq n)$, it does not represent $17m+16$; the same fact is true of the function $(1,2,2,2,3,5,a_7,\cdots,a_n)$ if $a_k\neq 6$

$(7 \leqq k \leqq n)$. It has been proved by direct verification that each of the remaining functions, that is, each function which satisfies the conditions of Theorem 3 and which represents $3m + 5, \cdots, 17m + 16$, does indeed represent $17m + 17, \cdots, 147m + 37$. These facts and Theorem 3 prove Theorem 4.

THEOREM 4. *If* $w = m + 4 \geqq 7$ *and if* $f = (1, 2, a_3, \cdots, a_n)$, *then* $f = 0, \cdots, 147m + 37$ *if, and only if,* f *satisfies* (5) *and* (7), *or* (6) *and* (7):

(5) $f = (1, 2, 2, 2, a_5, \cdots, a_n)$, $2 = a_4 = \cdots = a_J$, $a_{J+1} = 3$, $J = n - 1$ *or* $a_k \leqq w_{k-1} - 3$ $(J + 2 \leqq k \leqq n)$, *but* f *is not* $(1, 2, 2, 2, 3, 7, \cdots, a_n)$ *with* $n \geqq 7$ *and* f *is not* $(1, 2, 2, 3, 5, \cdots, a_n)$ *with* $n \geqq 7$ *and* $a_k \neq 6$ $(7 \leqq k \leqq n)$.

(6) $f = (1, 2, 2, 3, a_5, \cdots, a_n)$, $3 = a_4 = \cdots = a_{S-1}$, $a_S = 4$, $S = n$ *or* $a_k \leqq w_{k-1} - 3$ $(S + 1 \leqq k \leqq n)$,

(7) *if there is a coefficient* a_k *such that* $a_k = w_{k-1} - 3$ *then there is a coefficient greater than* a_k *and the first such coefficient is indeed* $a_k + 1$.

It has been proved by direct verification that: if f satisfies (1) then $f = 6m + 9, \cdots, 202m + 28$; if f satisfies (2) then $f = 6m + 9, \cdots, 387m + 116$; if f satisfies (3) or (4) then $f = 6m + 9, \cdots, 145m + 25$.

3. Universality of the functions in Theorem 2 and of the functions in Theorem 4.

Theorem 5 has been established by methods similar to those used in the proofs of the corresponding facts in I. Use was made of facts on simultaneous representation due to Dickson,[2] of results equivalent to his for the case $(1, 2, 2, 2)$ stated in Lemma 9, and of new results for the case $(1, 2, 2, 3)$ with b divisible by four stated in Lemma 8.

THEOREM 5. *If* f *satisfies the conditions of Theorem* 2 *or the conditions of Theorem* 4, *then there is a positive integer* M, *depending only on* m *and* f, *such that* f *represents every integer which is greater than* M. *If* f *satisfies* (1) *with* $a_5 = 1$, *then* $M = 44m + 40$; *if* f *satisfies* (1) *with* $a_5 = 2$, *or if* f *is* $(1, 1, 1, 1, 3)$, *then* $M = 142m + 76$; *if* f *satisfies* (2) *then* $M = 387m + 116$; *if* f *is* $(1, 1, 2, 3)$ *then* $M = 73m + 90$. *If* f *satisfies* (3) *and the value stated in Lemma* 7 *is used for* d, *then* $M = [11d(d - 5) + 74]m + 2(11d - 29)$; *if* f *satisfies* (5) *and* (7), *then* $M = [13d(d - 5) + 87]m + 2(13d - 379)$; *if* f *satisfies* (6) *and* (7), *then* $M = [15d(d - 5) + 100]m + 2(15d - 55)$.

LEMMA 7. *If* f *satisfies* (3) *then for each function the corresponding value of* d *appears in the following list:* $d = 2a_5 + 4$ *if* $a_5 = 3$ *or* 4; $d = 8$

if $a_5 = 2$ *and* $n = 5$; $d = 12$ *if* $a_5 = 2$ *and* $a_6 = 2$; $d = 2a_6 + 4$ *if* $a_5 = 2$ *and* $a_6 \geqq a_5 + 1$. *If f satisfies* (5) *with* $a_5 = 2$, *or with* $a_5 = 3$ *and* $a_6 = 5$, *then* $d = 9$; *if f satisfies* (5) *with* $a_5 = 3$, *and if* $n = 5$ *or* $a_6 = 3$ *or* $a_6 = 6$, *then* $d = 10$; *if f satisfies* (5) *with* $a_5 = 3$, *and if* $a_6 = 4$ *or* $a_6 = 7$, *then* $d = 12$. *If f satisfies* (6), *then for each function the corresponding value of d appears in the following list:* $d = 10$ *if* $a_5 = 3$; $d = 24$, $a_5 = 4$ *and* $n = 5$; $d = 2a_6 + 4$ *if* $a_5 = 4$ *and* $a_6 = 5$, 7, 9, *or* 6; $d = 28$ *if* $a_5 = 4$ *and* $a_6 = 8$, *or if* $a_5 = 4 = a_6$ *and* $n = 6$, *or if* $a_5 = 4 = a_6$ *and* $a_7 = 4$, 8, *or* 12; $d = 20$ *if* $a_5 = 4 = a_6$ *and* $a_7 = 5$, 6, *or* 7; $d = 2a_7 + 4$ *if* $a_5 = 4 = a_6$ *and* $a_7 = 9$, 10, 11, or 13.

Theorem 6 follows from Theorems 2, 4, and 5, and the facts stated at the end of Section 2.

THEOREM 6. *If f satisfies* (1), *or* (2), *or* (4), *then f represents every positive integer. If f satisfies* (3) *then f represents every positive integer A except perhaps when* $145m + 25 < A < M$. *If f satisfies* (5) *and* (7), *or if f satisfies* (6) *and* (7), *then f represents every positive integer except perhaps when* $147m + 37 < A < M$. *The values of M are given in Theorem* 5.

The universality of a function of weight $m + 4$ is implied by that of a function of weight $m + 3$ if and only if the coefficients of the former are obtained from the coefficients of the latter by prefixing a coefficient 1. The universality of a function of weight $m + 4$ is implied by that of a function of weight $m + 2$ if and only if the coefficients of the former are obtained from those of the latter by inserting a coefficient 2 or by prefixing two coefficients each equal to 1. It is thus proved, by comparison of the functions of Theorems 2 and 4 with the universal functions listed in I and II, that the universality of a function of weight $m + 4$ is implied by that of a function of weight $m + 2$ or that of a function of weight $m + 3$ if and only if f satisfies one of the conditions in the following list: f satisfies (2) with $a_5 = 2$, and with $n = 5$ or $a_k \leqq w_{k-1} - 3$ $(6 \leqq k \leqq n)$; f satisfies (1) with $a_5 = 2$ and $a_k \leqq w_{k-1} - 2$ $(6 \leqq k \leqq n)$; f satisfies (1) with $a_5 = 1$, and either $a_6 = 3$ and $n = 6$, or $a_6 = 2$ and $n = 6$ or $a_k \leqq w_{k-1} - 2$ $(7 \leqq k \leqq n)$, or $a_6 = 1$ and $a_k \leqq w_{k-1} - 2$ $(7 \leqq k \leqq n)$. These facts prove Theorem 7.

THEOREM 7. *The universality of a function of weight* $m + 4$ *is not implied by that of a function of weight* $m + 2$ *and is not implied by that of a function of weight* $m + 3$ *if and only if the function satisfies* (3), *or* (4), *or* (5) *and* (7), *or* (6) *and* (7), *or it satisfies* (1) *or* (2) *and is not in the preceding list.*

If $m \geqq 9$ the function f defined as follows satisfies Theorem 7. Define q

and r by $m-1=4q+r$, with $0 \leqq r \leqq 3$, and let $a_1=1=a_2=\cdots=a_5$, $a_6=4=\cdots=a_{q+4}$, $a_{q+5}=r+4$. Then f satisfies (1) with $a_5=1$ and $a_6=4$, and hence f is non-trivially universal. If $m=3,\cdots,8$ the following functions, respectively, are universal and satisfy Theorem 7: $(1,1,2,3)$, $(1,1,1,2,3)$, $(1,1,1,1,1,4)$, $(1,1,2,2,4)$, $(1,1,1,1,1,3,3)$, $(1,1,1,1,1,3,4)$.

4. Lemmas on simultaneous representation for the case (1, 2, 2, 3) with *b* divisible by four and for the case (1, 2, 2, 2). Dickson[2] proved that if a and b are integers such that

$$(8) \qquad a \equiv b \pmod 2, \qquad b^2 \leqq 8a, \qquad 7a \leqq b^2+2b+7, \qquad b \geqq 0,$$

and if b is not a multiple of four, then there are integers x, y, z, and w, such that each is non-negative and that

$$(9) \qquad a=x^2+2y^2+2z^2+3w^2, \qquad b=x+2y+2z+3w.$$

The case in which b is a multiple of four is treated in Lemma 8.

LEMMA 8. *If a and b are integers such that* (8) *holds, and if i and h are integers such that $i \geqq 2$ and $h \geqq 1$, and if*

$$(10) \qquad b=2^iB, \qquad a=a^{2h-1}A \text{ or } 2^{2h}A, \qquad A \text{ and } B \text{ odd},$$

then (9) *have a solution in integers, each non-negative, if and only if no one of the conditions* (11), (12), (13), (14), (15), *is satisfied:*

$$(11) \qquad a=2^{2h}A \text{ and } i=h+1, \text{ and } A \equiv 3 \ (\mathit{mod}\ 4),$$

$$(12) \qquad a=2^{2h-1}A \text{ and } i=h+1 \text{ and } A-B^2 \text{ of the form } 4^t(8s+5),$$

$$(13) \qquad a=2^{2h-1}A \text{ and } i \geqq h+3 \text{ and } A \equiv 5 \ (\mathit{mod}\ 8),$$

$$(14) \qquad a=2^{2h-1}A \text{ and } i=h+2 \text{ and } A \equiv 1 \ (\mathit{mod}\ 8),$$

$$(15) \qquad a=2^{2h-1}A \text{ and } i>h=1 \text{ and } A \equiv 1 \ (\mathit{mod}\ 4).$$

The proof of this lemma is based on the fact, which Dickson proved in Section 4 of the paper to which reference has been made, that there are integers x, y, z, w satisfying (9) if and only if there are integers V, D, W such that

$$(16) \qquad 8a-b^2=V^2+8D^2+6W^2,$$

$$(17) \qquad b+V-4D \equiv 0 \pmod 8, \qquad b-2D-W \equiv 0 \pmod 4.$$

The proof also uses the fact that if N is an odd integer then there are integers u and v which are both odd such that $4N=u^2+3v^2$ if, and only if, there are integers u_1 and v_1 which are both even such that $4N = u_1^2 + 3v_1^2$. Hence, by using (10) in (16) and an exhaustive separation into cases, it is proved that, if there are integers V, D, W satisfying (16), then there are integers x, y, z, w satisfying (9) if and only if (15) does not hold. Now $8a - b^2$ is a multiple of four, and a multiple of four is represented by the form $V^2 + 8D^2 + 6W^2$ if and only if[3] it is not of the form $4^t(8s + 5)$. By using (10) it is proved that $8a - b^2 = 4^t(8s + 5)$ if and only if one of the conditions (11)-(14) is satisfied. All these facts prove Lemma 8.

Dickson proved that if a and b are integers such that

$$(18) \qquad a \equiv b \pmod 2, \qquad b^2 \leqq 7a, \qquad 6a \leqq b^2 + 2b + 6, \qquad b \geqq 0,$$

and if $7a - b^2$ is not of the form $4^t(16s + 14)$, then there are integers x, y, z, w, such that each is non-negative and that

$$(19) \qquad a = x^2 + 2y^2 + 2z^2 + 2w^2, \qquad b = x + 2y + 2z + 2w.$$

If the notations (10) are used in the equation $7a - b^2 = 4^t(16s + 14)$, the conditions stated in Lemma 9 are verified.

LEMMA 9. *If a and b are integers such that* (18) *holds, and if i and h are integers such that $i \geqq 0$ and $h \geqq 0$ and the notations* (10) *are used, then* (19) *have a solution in integers, each non-negative, if and only if no one of conditions* (20), (21), (22), (23) *is satisfied:*

(20) $a = 2^{2h}A$ *and* $i = h \geqq 0$, *and* $A \equiv 1 \pmod{16}$ *and* $B \equiv \pm 3 \pmod 8$,

(21) $a = 2^{2h}A$ *and* $i = h \geqq 0$, *and* $A \equiv 9 \pmod{16}$ *and* $B \equiv \pm 1 \pmod 8$,

(22) $a = 2^{2h-1}A$ *and* $i = h \geqq 1$, *and* $A \equiv 3 \pmod 8$,

(23) $a = 2^{2h-1}A$ *and* $i \geqq h + 1 \geqq 2$, *and* $A \equiv 1 \pmod 8$.

NORTHWESTERN UNIVERSITY.

---

[3] B. W. Jones, *University of Chicago Dissertation* (1928), p. 122.

THE HOMOMORPHISM THEOREMS FOR LOOPS.*

By REINHOLD BAER.

A loop is a system L of elements with one composition, called addition, $x+y$, subject to the following rules:

I. If two of the three elements x, y, z are given, then the equation $x+y=z$ has one and only one solution in L.

II. There exists a nul-element 0, satisfying $x+0=0+x=x$.

Thus addition, right- and left-subtraction are unique; and this makes the term subloop self-explanatory.

A homomorphism is a single-valued and addition-preserving mapping of one loop into (the same or another) loop. It has been noted that in the theory of homomorphisms classical group-theory does not make much use of the associative law; and the present note is intended to contribute towards a substantiation of this contention.

The set of all elements in a loop which are mapped upon 0 by a given homomorphism is a subloop which is termed the kernel of this homomorphism; and every subloop which is the kernel of some homomorphism[1] has been called a normal subloop. The following theorems are due to A. A. Albert and M. F. Smiley.[2]

The subloop S of the loop L is a normal subloop if, and only if,

$$\begin{gathered} x+S=S+x, \qquad (x+y)+S=x+(y+S), \\ (x+S)+y=x+(S+y), \qquad (S+x)+y=S+(x+y) \end{gathered} \tag{1}$$

for x and y in L.

If S is a normal subloop of the loop L, then the sets $S+x$ for x in L are termed the cosets of L modulo S. These cosets form the quotient loop L/S; and mapping the element x in L upon the coset $S+x$ constitutes the natural homomorphism of L upon L/S.

The isomorphism theorems for quotient groups including Zassenhaus' four subgroups formula have been proved for loops too.[3] But it seems to the

---

* Received December 6, 1944.

[1] Most writers on loop theory define normality by this or an equivalent property.

[2] Cf. Smiley (1), Theorem 2, p. 783.

[3] Albert (2), Theorems 4 and 5, pp. 404-405; Smiley (1), Theorem 5, p. 784.

author that Zassenhaus' proof[4] is still the simplest and most advantageous one, though it has, apparently, not yet been used for loops. This proof is offered here for two reasons: firstly some slight changes are necessitated through omission of the associative law; and thus it may be convenient to have available a detailed exposition. Secondly it forms the basis for a deeper investigation of these laws; and we are able to prove both extensions and converses of them which are new in the associative case too.

The distinctive feature of the method used here is a preference for the use of homomorphisms instead of computations with elements. Thus of the two definitions of normal subloops mentioned above we neglect the one which expresses itself in the equations (1) and we utilize mainly the characterization as the kernel of a homomorphism.

The isomorphism laws are known to lead immediately to a proof of the Jordan-Hölder-Schreier-Zassenhaus Theorem. We give an indication of this proof too, since the frame work given is slightly more general than the one customarily used, and since in the case of loops there occur some possibilities beyond those encountered in the classical case.

Notations. If S and T are subsets of the loop L, then we denote by $S+T$ the *set* of all the elements $s+t$ for s in S and t in T; by $\{S,T\}$ the subloop generated by S and T, i.e., the smallest subloop of L containing S and T; and by $S \cap T$ the cross cut of S and T.

If x is an element in the loop L, and if η is a homomorphism of L, the image of x under η is denoted by $x\eta$; and likewise we denote by $S\eta$ the image of the subset S of L. We say that η is a homomorphism of L *into* H, if $L\eta$ is a subset, and therefore a subloop, of the loop H; and if $L\eta = H$, then η is termed a homomorphism of L *upon* H. As usual an isomorphism is a homomorphism which maps different elements upon different elements.

1. The first isomorphism theorem.

THEOREM 1. *If S and T are subloops of the loop L, and if S is a normal subloop of $\{S,T\}$, then*

(a) $\{S,T\} = S+T = T+S$;

(b) *$S \cap T$ is a normal subloop of T;*

(c) *reciprocal isomorphisms between $\{S,T\}/S$ and $T/(S \cap T)$ are obtained by mapping the coset X of $\{S,T\}/S$ upon the set $X \cap T$ and by mapping the coset Y of $T/(S \cap T)$ upon the set $S+Y$.*

---

[4] Zassenhaus (1), pp. 32-33 and pp. 50-52. Of course, Zassenhaus is concerned only with the associative case.

Proof. (a) may either be verified by reference to the characteristic equations (1), stated in the introduction, or by the following argument. Since S is a normal subloop of the loop $K = \{S, T\}$, there exists a homomorphism η of K (upon some loop J) whose kernel is exactly S. Clearly $K\eta = \{S\eta, T\eta\} = T\eta$. Hence there exists to every element x in K an element x^* in T such that $x\eta = x^*\eta$. Consequently x and x^* belong[5] to the same coset of K modulo S so that x belongs to $S + x^* = x^* + S$. Since x^* is an element in T, we have found that $K = S + T = T + S$.

The homomorphism η of K defines a homomorphism of the subloop T of K (upon the subloop $T\eta$ of H). The kernel of this homomorphism of T is obviously the cross cut $S \cap T$ of T and of the kernel S of η. Thus $S \cap T$ is a normal subloop of T.

We have noted before that $K\eta = T\eta = J$ so that $\{S, T\}$ and T have the same image under the homomorphism η. If y is any element in J, then denote by y' the set of all the elements in K that are mapped by η upon y; and denote by y'' the set of all the elements in T that are mapped by η upon y. Clearly $y'' = T \cap y'$ and $y' = S + y''$; and (c) is an immediate consequence of these facts.

LEMMA. *If N is a normal subloop of the loop L, if η is a homomorphism of L, and if $N\eta = 0$, then η defines a homomorphism of L/N.*

Proof. If x and y are any two elements in L, then there exists one and only one element z in L such that $z + y = x$. If x and y belong to the same coset modulo N, then z is in N (and conversely). In this case $y\eta = z\eta + y\eta = (z + y)\eta = x\eta$; and our contention is an immediate consequence of this fact.

THEOREM 2. *If M and N are normal subloops of the loops K and L, respectively, and if η is a homomorphism of K into L such that $M\eta \leq N$, then a homomorphism of K/M into L/N is defined by mapping the coset X of K/M upon $N + X\eta$.*

Proof. Denote by ν the natural homomorphism of L upon L/N which maps the element y in L upon the coset $N + y$. Since the product of two homomorphisms is a homomorphism, if defined, $\eta\nu$ is a homomorphism of K into L/N. From $M\eta \leqq N$ we infer $M\eta\nu = 0$; and hence it follows from Lemma 1 that $\eta\nu$ defines a homomorphism of K/M into L/N. Clearly $X\eta\nu = N + X\eta$ for X in K/M, completing the proof.

---

[5] The uniquely determined solutions y, z of the equations $y + x^* = x$ and $x^* + z = x$ belong to S.

Theorem 2 leads us to the following convenient *definition*: If M and N are normal subloops of the loops K and L respectively, and if η is a homomorphism of K into L such that $M\eta \leq N$, then *the homomorphism of* K/M *into* L/N *which is induced by* η maps X in K/M upon $N + X\eta$. It is clear that η cannot induce a homomorphism of K/M into L/N, unless $M\eta \leq N$; and thus the statement: "η induces a homomorphism of K/M into L/N" shall always be understood to imply $M\eta \leq N$.

It may be worth mentioning that the three Isomorphism Theorems are either contained in or may be inferred from the results obtained so far.

THEOREM 3. *The following properties of the subloops* R *and* S *of the loop* L *and of the normal subloops* M *and* N *of* R *and* S*, respectively, imply each other.*

(i) $N = M \cap S$ *and* $R = M + S$.

(ii) *The identity automorphism of* L *induces an isomorphism of* S/N *upon* R/M.

(iii) $N \leq M$, $S \leq R$; *and there exists a homomorphism* η *of* L, *a subloop* T *of* $L\eta$ *and a normal subloop* P *of* T *such that* η *induces both an isomorphism of* R/M *upon* T/P *and an isomorphism of* S/N *upon* T/P.

Proof. It is an immediate consequence of Theorem 1 (c) that (i) implies (ii). If (ii) holds true, then let $\eta = 1$, $R = T$, $M = P$; and it is evident that (ii) implies (iii).

Assume finally the validity of (iii). Denote by A the set of all the elements in L which are mapped by η upon elements in T; and denote by B the set of all the elements in L which are mapped by η upon elements in P. Clearly A and B are subloops of L. Mapping x in A upon the coset $P + x\eta$ in T/P we obtain a homomorphism of A into T/P whose kernel is B, showing that B is a normal subloop of A. Since η is supposed to induce an isomorphism of R/M upon T/P, we have $N \leq M \leq B$ and $S \leq R \leq A$. The only elements in R which are mapped by η upon elements P are the elements in M, implying $M = B \cap R$; and likewise one deduces $N = B \cap S$ from the fact that η induces an isomorphism of S/N upon T/P. Consequently we have:

$$N \leq M \cap S \leq B \cap S = N \quad \text{or} \quad N = M \cap S.$$

If r is an element in R, then $P + r\eta$ is a coset of T/P, since η induces an isomorphism of R/M upon T/P. But η induces an isomorphism of S/N upon T/P; and hence there exists (Theorem 2) an element s in S such that $P + r\eta = P + s\eta$. Consequently there exists an element p in P such that

$r\eta = p + s\eta$. There exists furthermore one and only one element q in R such that $r = q + s$, since r and s are both in the subloop R of L. Hence

$$p + s\eta = r\eta = (q + s)\eta = q\eta + s\eta, \quad \text{proving} \quad p = q\eta.$$

Hence q belongs to the cross cut $B \cap R = M$; and $r = q + s$ belongs to $M + S$. Thus we have shown that $M + S \leq R \leq M + S$ or $R = M + S$, proving that (i) is a consequence of (iii).

Corollary 1. *If M and S are normal subloops of the subloop R of the loop L, if N is a normal subloop of S, and if conditions (i) to (iii) of Theorem 3 are satisfied by M, N, R, S, then N is a normal subloop of M and M/N and R/S are isomorphic loops.*

This is an immediate consequence of Theorem 1 and of the symmetry of condition (i).

Theorem 4. *The following properties of the subloops R and S of the loop L and of the normal subloops M and N of R and S, respectively, imply each other.*

(i) *$N \leq M$ and $R = M + S$.*

(ii) *The identity automorphism of L induces a homomorphism of S/N upon R/M.*

(iii) *$N \leq M$, $S \leq R$; and there exists a homomorphism η of L, a subloop T of $L\eta$ and a normal subloop P of T such that η induces an isomorphism of R/M upon T/P and a homomorphism of S/N upon T/P.*

Proof. If (i) holds true, then mapping the coset X of S/N upon $M + X$ constitutes the homomorphism of S/N upon R/M which is induced by the identity automorphism of L, proving that (ii) is a consequence of (i).

To prove that (iii) is a consequence of (ii), one puts $\eta = 1$, $R = T$ and $M = P$.

If finally (iii) is satisfied, then denote by E the set of all the elements in S which are mapped by η upon elements in P. It is readily seen that $N \leq E$ and that E/N is the kernel of the homomorphism of S/N upon T/P which is induced by η. Thus E is the kernel of the homomorphism $\nu\eta$ for ν the natural homomorphism of S upon S/N. Since E is part of R, and since η induces an isomorphism of R/M upon T/P, we infer $E \leq M$ from $E\eta \leq P$. Since η induces clearly an isomorphism of S/E upon T/P, we may apply Theorem 3, (iii) proving that $E = M \cap S$ and $R = M + S$. Thus we have shown that (i) is a consequence of (iii), and that furthermore the following statement holds true.

COROLLARY 2. *If the subloops M, N, R, S meet the requirements of Theorem 4, and if condition (iii) of Theorem 4 is satisfied by them, then* $(M \cap S)/N$ *is the kernel of the homomorphism of* S/N *upon* T/P *which is induced by* η.

2. Zassenhaus' isomorphism theorem.

LEMMA 1. *If N is a normal subloop of the loop L, then the homomorphism η of L maps N upon a normal subloop $N\eta$ of $L\eta$.*

This is an immediate consequence of the formulas (1), characteristic for normal subloops, which we stated in the introduction.

LEMMA 2.[6] *If U and V are subloops of the loop L, and if W is a subloop of U, then*[7] $(U \cap V) + W = U \cap (V + W)$.

Proof. Since $W \leq U$ and $V \leq V + W$, we have clearly $(U \cap V) + W \leqq U \cap (V + W)$. If x is an element in $U \cap (V + W)$, then x belongs to U and there exist elements v and w in V and W, respectively, such that $x = v + w$. But w belongs, as an element in W, to U; and x belongs to U. Since U is a subloop of L, this implies that v belongs to U too so that v is in $U \cap V$ and x is in $(U \cap V) + W$, proving $U \cap (V + W) \leq (U \cap V) + W$; and this completes the proof of the desired equation.

THEOREM 1. *If R and S are subloops of the loop L, and if M and N are normal subloops of R and S respectively, then*

(a) $(R \cap N) + M$ *is a normal subloop of the subloop* $(R \cap S) + M$ *of* L,

(b) $(M \cap S) + (N \cap R)$ *is a normal subloop of* $R \cap S$,

(c) *the identity automorphism of L induces an isomorphism of*

$$(R \cap S)/[(M \cap S) + (N \cap R)] \quad \textit{upon} \quad [(R \cap S) + M]/[(R \cap N) + M].$$

Proof.[8] Since M is a normal subloop of R, it is a normal subloop of both $\{(R \cap N), M\}$ and $\{(R \cap S), M\}$, as follows from Theorem 1, (b) of **1**.

---

[6] Ore's quadrilateral condition is an almost immediate consequence of this Lemma 2 and of Theorem 1 of 1; see Ore (1) and Smiley (1), Theorem 5 on p. 784.

[7] This equation is an equality of sets of elements in a loop, not necessarily of subloops. The author is indebted to Mr. R. Good for this improvement of his original version.

[8] The proof is an adaptation of the proof by Zassenhaus (1), p. 50 *et seq.*

Hence it follows from Theorem 1, (a) of **1** that

$$\{(R \cap N), M\} = (R \cap N) + M \quad \text{and} \quad \{(R \cap S), M\} = (R \cap S) + M.$$

Since N is a normal subloop of S, there exists a homomorphism η of S whose kernel is exactly N. Clearly η defines a homomorphism on the subloop $R \cap S$ of S whose kernel is $N \cap R$. Thus $N \cap R$ is a normal subloop of $R \cap S$. The natural homomorphism of R upon R/M maps $R \cap N$ upon $[(R \cap N) + M]/M$; and $R \cap S$ upon $[(R \cap S) + M]/M$; and hence it follows from Lemma 1 that $[(R \cap N) + M]/M$ is a normal subloop of $[(R \cap S) + M]/M$. Consequently there exists a homomorphism v of $[(R \cap S) + M]/M$ whose kernel is just $[(R \cap N) + M]/M$. If ν is the natural homomorphism of $(R \cap S) + M$ upon $[(R \cap S) + M]/M$, then νv is a homomorphism whose kernel is exactly $(R \cap N) + M$, showing the validity of (a).

It is a consequence of Lemma 2 that

$$(R \cap S) \cap [M + (R \cap N)] = (R \cap S \cap M) + (R \cap N) = (S \cap M) + (R \cap N).$$

It is a consequence of (a) that $(R \cap N) + M$ is a normal subloop of the subloop $(R \cap S) + M$ of L. Hence it follows from Theorem 1, (b) of **1** that $(R \cap S) \cap [(R \cap N) + M]$ is a normal subloop of $R \cap S$, since

$$(R \cap S) + M = (R \cap S) + [(R \cap N) + M].$$

This proves our present contention (b). Finally (c) is an immediate consequence of the results already obtained in conjunction with Theorem 1, (c) (or Theorem 3) of **1**.

COROLLARY 1. *If R and S are subloops of the loop L, and if M and N are normal subloops of R and S, respectively, then*

$$[(R \cap S) + M]/[(R \cap N) + M] \quad \textit{and} \quad [(R \cap S) + N]/[(S \cap M) + N]$$

are isomorphic quotient loops.

This is an immediate consequence of Theorem 1, (c) and of the symmetry of the hypotheses.

THEOREM 2. *The following properties of the subloops R and S of the loop L and of the normal subloops M and N of R and S, respectively, imply each other.*

(i) $R = (R \cap S) + M$ *and* $S = (R \cap S) + N$.

(ii) *There exists a subloop T of L such that the identity automorphism of L induces both a homomorphism of T upon R/M and a homomorphism of T upon S/N.*

(iii) *There exists a homomorphism* η *of a loop* H *into* L *which induces both a homomorphism of* H *upon* R/M *and a homomorphism of* H *upon* S/N.

Proof. If (i) is satisfied by R, S, M, N, then we put $T = R \cap S$. Since $0 \leq M$ and $R = T + M$, it follows from Theorem 4 of **1** that the identity automorphism of L induces a homomorphism of T upon R/M; and likewise we see that the identity automorphism of L induces a homomorphism of T upon S/N, proving that (ii) is a consequence of (i).

If (ii) holds, then let $H = T$ and denote by η the homomorphism of H into L which is obtained by mapping every element in T upon itself. Thus (iii) is a consequence of (ii).

Assume finally the validity of (iii). Then $R = \{M, H\eta\}$. But M is a normal subloop of R; and thus it follows from Theorem 1, (a) of **1** that $R = M + H\eta$. Clearly $H\eta \leq S$; and thus we find that

$$R = M + H\eta \leq M + (R \cap S) \leq M + R = R \quad \text{or} \quad R = M + (R \cap S);$$

and likewise we verify $S = N + (R \cap S)$, proving that (i) is a consequence of (iii).

THEOREM 3. *The following properties of the subloops* R *and* S *of the loop* L *and of the normal subloops* M *and* N *of* R *and* S, *respectively, imply each other.*

(i) *There exists a subloop* T *of* L *and a normal subloop* P *of* T *such that the identity automorphism of* L *induces both an isomorphism of* T/P *upon* R/M *and an isomorphism of* T/P *upon* S/N.

(ii) *There exists a normal subloop* Q *of a loop* H *and a homomorphism* η *of* H *into* L *which induces both an isomorphism of* H/Q *upon* R/M *and an isomorphism of* H/Q *upon* S/N.

Proof. If (i) is true then the validity of (ii) is shown by taking $H = T$, $Q = P$ and η that homomorphism of H into L which maps every element in T upon itself.

If conversely (ii) is satisfied by R, S, M, N, then we put $T = H\eta$ and $P = Q\eta$. It is clear that T and P are subloops of L; and it is a consequence of Lemma 1 that P is a normal subloop of T. Since η induces an isomorphism α of H/Q upon R/M, we have $T \leq R$ and $P \leq M$. Thus the identity automorphism of L induces a homomorphism β of T/P into R/M. We note finally that η induces a homomorphism γ of H/Q upon T/P. Consequently $\alpha = \gamma\beta$; and β must, therefore, be an isomorphism of T/P upon R/M; and likewise we see that the identity automorphism of L induces an isomorphism of T/P upon S/N, proving that (i) is a consequence of (ii).

THEOREM 4. *The following properties of the subloops R and S of the loop L and of the normal subloops M and N of R and S, respectively, imply each other.*

(i) $R = (R \cap S) + M, S = (R \cap S) + N,$
$M = (R \cap N) + M, N = (S \cap M) + N.$

(ii) $R \cap N = S \cap M$; *and there exists a normal subloop P of a subloop T of L such that the identity automorphism of L induces an isomorphism of T/P upon R/M and an isomorphism of T/P upon S/N.*

Proof. If (i) is satisfied, then $R \cap N \leq M$ and hence $M \cap N \leq R \cap N \leq M \cap N$ or $R \cap N = M \cap N$; and likewise we find that $S \cap M = M \cap N$, proving the first part of condition (ii). Let $T = R \cap S$ and $P = M \cap N$. Then $P = (R \cap N) + (S \cap M)$; and it follows from Theorem 1, (b) that P is a normal subloop of T. From Theorem 1, (c) we infer that the identity automorphism induces an isomorphism of T/P upon R/M, since—by (i)—$R = (R \cap S) + M$ and $M = (R \cap N) + M$; and likewise we see that the identity automorphism of L induces an isomorphism of T/P upon S/N, proving that (ii) is a consequence of (i).

Assume, conversely, the validity of (ii). Then we infer from Theorem 3 of **1** that

$$R = M + T, \quad P = M \cap T \quad \text{and} \quad S = N + T, \quad P = N \cap T.$$

As in the proof of Theorem 2 we see that

$$R = M + T \leq M + (R \cap S) \leq R \quad \text{or} \quad R = M + (R \cap S)$$

and likewise we verify $S = N + (R \cap S)$. From the first part of condition (ii) we deduce that

$$M \leq M + (R \cap N) = M + (S \cap M) \leq M \quad \text{or} \quad M = M + (R \cap N);$$

and in the same fashion we show that $N = N + (S \cap M)$, proving that (i) is a consequence of (ii).

Remark. The impossibility of omitting the first part of condition (ii) may be seen from the following considerations. Suppose that L is an abelian group (= an associative and commutative loop), and that the subgroups M, N, T satisfy: $M \cap T = N \cap T = 0$. Clearly $R = M + T$ is the direct sum of M and T; and $S = N + T$ is the direct sum of N and T. Thus the identity automorphism of L induces an isomorphism of T upon R/M and of T upon

S/N. But clearly $R \cap N$ and $S \cap M$ need not be equal. As an example let L be a free abelian group of rank 2; and let t, m be a basis of L. Put $T = \{t\}$, $M = \{m\}$ and $N = \{m + t\}$. Then $T \cap M = T \cap N = 0$, but

$$R \cap N = (M + T) \cap N = L \cap N = N, \qquad S \cap M = (N + T) \cap M = L \cap M = M \neq N.$$

3. The Theorems of Jordan–Hölder–Schreier–Zassenhaus. Suppose that M is a system of multipliers ($=$ operators) for the elements in the loop L. Then there exists to every element x in L and to every element m in M a uniquely determined element xm in L, subject to the distributive law:

$$(x + y)m = xm + ym \text{ for } x, y \text{ in } L \text{ and } m \text{ in } M.$$

Clearly every element in M induces an endomorphism of L ($=$ homomorphism of L into itself).

A subloop S of L is *M-admissible,* if $SM \leq S$. If S and T are M-admissible subloops of L, and if S is a normal subloop of T, then we infer from the lemma of **1** that M may be considered as a system of multipliers for T/S too.

Suppose now that the system Δ of subloops of L meets the following requirements.

(a) *0 and L belong to Δ.*

(b) *If S and T belong to Δ, then $S \cap T$ and $\{S, T\}$ belong to Δ.*

(c) *Every subloop in Δ is M-admissible.*

The subloops in Δ will be referred to as the *Δ-subloops of L.*

Examples of systems Δ: The system of all M-admissible subloops; the system of all M-admissible normal subloops of L; the system of all M-admissible and M-characteristic subloops of L; the system of all M-admissible and M-characteristic normal subloops of L etc.

Here we say that an endomorphism ϵ of L is an M-endomorphism, if $xm\epsilon = x\epsilon m$ for every x in L and every m in M; and terms like M-characteristic, M-fully-invariant etc. are now self-explanatory. It should be noted, however, that fully invariant subloops need not be normal.[9]

A *Δ-chain of length n* is a system C of Δ-subloops $S(i)$ of L such that

$$S(0) = 0,\ S(i-1) \textit{ is a normal subloop of } S(i),\ S(n) = L.$$

---

[9] The author is indebted to Professor R. H. Bruck for pointing out this possibility. Bruck (1), Theorem 1, pp. 770-771 shows the existence of a simple loop H containing one and only one maximal subloop $M \neq 0$, $M \neq H$. Clearly M is not normal, but fully invariant.

If C and D are Δ-chains of length m and n respectively, then D is termed a refinement of C, if $m \leq n$ and C is a subset of D. The Δ-chains $C = (\cdots C(i) \cdots)$ and $D = (\cdots D(j) \cdots)$ are said to be isomorphic, if they are of equal length n, and if there exists a permutation: $i \rightarrow i'$ of the integers between 0 and n such that $C(i+1)/C(i)$ and $D(i'+1)/D(i')$ are M-isomorphic.

THEOREM. *Any two Δ-chains of L possess isomorphic refinements.*

Proof.[10] If $C(i)$ for $0 \leq i \leq m$ and $D(j)$ for $0 \leq j \leq n$ are two Δ-chains of L, then it follows from Theorem 1 of **2** that

$$[(C(i+1) \cap D(j+1)) + C(i)]/[(C(i+1) \cap D(j)) + C(i)]$$

and

$$[(C(i+1) \cap D(j+1)) + D(j)]/[(C(i) \cap D(j+1)) + D(j)]$$

are M-isomorphic loops. Letting $C^*(in+j) = (C(i+1) \cap D(j)) + C(i)$, $D^*(jm+i) = (C(i) \cap D(j+1)) + D(j)$ and $(in+j)' = jm+i$, the proof is easily completed.

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

BIBLIOGRAPHY

A. A. Albert:

(1) "Quasigroups I," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 507-519.

(2) "Quasigroups II," *Transactions of the American Mathematical Society*, vol. 55 (1944), pp. 401-419.

R. H. Bruck:

(1) "Simple Quasigroups," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 769-781.

R. J. Duffin and R. S. Pate:

(1) "An abstract theory of the Jordan-Hölder composition series," *Duke Mathematical Journal*, vol. 10 (1943), pp. 743-750.

O. Ore:

(1) "Chains in partially ordered sets," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 558-566.

M. F. Smiley:

(1) "An application of lattice theory to quasigroups," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 782-786.

H. Zassenhaus:

(1) *Lehrbuch der Gruppentheorie.* Erster Band. Hamburger Math. Einzelschriften No. 21 (1937).

---

[10] Cf. Zassenhaus (1), p. 51.

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

ON THE REPRESENTATION OF A GROUP OF ORDER *g* IN THE FIELD OF THE *g*-TH ROOTS OF UNITY.*

By RICHARD BRAUER.

To Hermann Weyl on his sixtieth birthday.

Introduction. It has long been surmised that every irreducible representation $\mathfrak{L}$ of a group $\mathfrak{G}$ of order g can be written in the field of the g-th roots of unity. This question has attracted the efforts of a number of mathematicians. Improving an earlier result of H. Maschke [**20**],[1] W. Burnside [**14**] proved that this conjecture is actually true for a very extensive class of representations. I. Schur [**22**] showed that it is correct for all soluble groups. The case of representations of odd degree with real character was treated by A. Speiser [**25**].[2] The most important progress after that was made by H. Hasse [**11**] who showed that the representation $\mathfrak{L}$ can be written in the field of the g^n-th roots of unity where n is a suitable positive rational integer depending on $\mathfrak{L}$.

In the present paper, the old question is solved in its full generality for the first time. We prove the theorem:

THEOREM. *If $\mathfrak{G}$ is a group of finite order g, then every irreducible representation $\mathfrak{L}$ of $\mathfrak{G}$ can be written in the field Ω of the g-th roots of unity* (*that is, $\mathfrak{L}$ is similar to a representation with coefficients in Ω*).

The proof is obtained by combining the methods of I. Schur and of H. Hasse with results concerning the modular representations of groups.[3]

1. Remarks on notation. In the following, $\mathfrak{G}$ will denote a group of finite order g. If G is a fixed element of $\mathfrak{G}$, the elements of $\mathfrak{G}$ commuting with G form a subgroup $\mathfrak{N}(G)$, the *normalizer* of G. The class of conjugate elements containing the element G then consists of $g/n(G)$ elements where $n(G)$ is the order of $\mathfrak{N}(G)$.

---

* Received August 19, 1945.

[1] Numbers in square brackets refer to the bibliography at the end.

[2] See also Schur [**24**], Brauer [**3**].

[3] For the definitions and theorems of this theory as far as they are used here, see Brauer-Nesbitt [**12**] and [**13**] Parts I and II, Nakayama, [**21**], Brauer [**7**], [**8**], [**9**], sections 1 and 4, [**10**], sections 1-4. Part I of [**13**] contains a short summary of the results and proofs of [**12**], with exception of the proof of Theorem I of [**12**]. Proofs of this theorem can also be found in [**7**], [**21**], and in Artin-Nesbitt-Thrall, [**2**], Chapter IX, 8.

A representation $\mathfrak{L}$ of $\mathfrak{G}$ is a group of matrices $\mathfrak{L}$ on which $\mathfrak{G}$ is mapped homomorphically; the coefficients of the matrices lie in a given field Λ. The image in $\mathfrak{L}$ of an element G of $\mathfrak{G}$ will be denoted by $\mathfrak{L}(G)$ and an analogous notation will be used throughout the paper. The word *representation* without further remark will always refer to a representation in a field of characteristic 0. In the other case, the term *modular representation* will be used. If p is the characteristic of the field in this case, then the elements R of $\mathfrak{G}$ whose order is prime to p are of importance; we denote them as the *p-regular elements* of $\mathfrak{G}$.

The group $\mathfrak{G}$ has k distinct absolutely irreducible[4] representations $\mathfrak{Z}_1, \mathfrak{Z}_2, \cdots, \mathfrak{Z}_k$ where k is the number of classes of conjugate elements in $\mathfrak{G}$. These representations can be chosen with coefficients in a suitable algebraic number field Λ of finite degree over the field P of rational numbers. If p is a given rational prime, and $\mathfrak{q}$ a prime ideal divisor of p in Λ, we may assume without restriction that all coefficients of every $\mathfrak{Z}_\kappa(G)$, G in $\mathfrak{G}$, are $\mathfrak{q}$-integers, that is, that they are quotients u/v of integers of Λ such that v is not divisible by $\mathfrak{q}$. If every coefficient of $\mathfrak{Z}_\kappa$ is replaced by its residue class modulo $\mathfrak{q}$, a modular representation $\mathfrak{Z}^*_\kappa$ in a field of characteristic p is obtained. The distinct constituents of $\mathfrak{Z}^*_1, \mathfrak{Z}^*_2, \cdots, \mathfrak{Z}^*_k$ furnish a complete system of absolutely irreducible modular representations $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_l$ for the characteristic p. The numbers $d_{\kappa\lambda}$ indicating how often $\mathfrak{F}_\lambda$ appears as a constituent in $\mathfrak{Z}^*_\kappa$ are the *decomposition numbers* (for p); they are non-negative rational integers.

The character of $\mathfrak{Z}_\kappa$ will be denoted by ζ_κ and the character of $\mathfrak{F}_\lambda$ by ϕ_λ.[5] We then have

$$\zeta_\kappa(R) = \sum_{\lambda=1}^{l} d_{\kappa\lambda}\phi_\lambda(R) \tag{1}$$

for p-regular elements R of $\mathfrak{G}$.

2. Proof of the theorem for modular representations $\mathfrak{L}$. If $\mathfrak{L}$ is an irreducible modular representation $\mathfrak{F}$ of $\mathfrak{G}$ in a field Π of characteristic p, then Π will certainly contain the traces $\mathrm{tr}(\mathfrak{F}(G))$ of all the matrices $\mathfrak{F}(G)$. Let Π_0 be the field generated by these g traces. Since every trace is a sum of

---

[4] A representation is absolutely irreducible, if it remains irreducible in every extension field.

[5] It should be observed that the modular characters are defined as complex numbers, not as elements of a modular field, see for instance [13], § 6. The trace of the matrix $\mathfrak{F}_\lambda(R)$ for a p-regular element R of $\mathfrak{G}$ then is the residue class $\phi_\lambda(R)^*$ of $\phi_\lambda(R)$ (mod $\mathfrak{q}$).

g-th roots of unity, the field Π_0 is a finite Galois field. It follows[6] that the representation $\mathfrak{F}$ can be written in the field Π_0 itself. As Π_0 is a subfield of the field of the g-th roots of unity, this proves the theorem for modular representations $\mathfrak{L}$.

3. A connection between the $\mathfrak{p}$-indices[7] and the decomposition numbers. We now come to the much more difficult case of non-modular representations $\mathfrak{L}$. It will be necessary to derive first a number of lemmas.

Let K_0 be the field obtained by adjoining all the modular characters $\phi_1(R), \phi_2(R), \cdots, \phi_l(R)$ (for all p-regular elements R of $\mathfrak{G}$) to the field of rational numbers P,

$$(2) \qquad K_0 = P(\phi_1, \phi_2, \cdots, \phi_l).$$

To K_0, adjoin the values $\zeta_\kappa(G)$ of a fixed character ζ_κ for all elements G of $\mathfrak{G}$,

$$(3) \qquad K = K_0(\zeta_\kappa).$$

Let p be a rational prime number and let $\mathfrak{p}$ be a fixed prime ideal of K dividing p. If θ is any $\mathfrak{p}$-integer of K we denote its residue class $(\mathrm{mod}\ \mathfrak{p})$ by θ^*. The totality of all θ^* forms the residue class field K^* of K $(\mathrm{mod}\ \mathfrak{p})$.

If R is a p-regular element of $\mathfrak{G}$, the trace of $\mathfrak{F}_\lambda(R)$ is equal to $\phi_\lambda(R)^*$.[8] If G is an arbitrary element of $\mathfrak{G}$, it may be written in the form $G = PR$ where the order of P is a power of p while R is a p-regular element belonging to $\mathfrak{N}(P)$. Then[9]

$$\mathrm{tr}(\mathfrak{F}_\lambda(G)) = \mathrm{tr}(\mathfrak{F}_\lambda(R)) = \phi_\lambda(R)^*.$$

In any case, the field K^* contains the traces of all matrices $\mathfrak{F}_\lambda(\mathfrak{G})$. The result of Section 2 shows that $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_l$ may be written with coefficients in K^*.

It has been shown previously[10] that there exists a representation $\mathfrak{W}_\lambda$ of $\mathfrak{G}$

---

[6] Brauer [5] §4, p. 101. The fact used here is equivalent to the theorem of Wedderburn [27] that every division algebra over a finite field is commutative.

[7] The theory of the index of a representation with regard to a field was developed by I. Schur, [22], [23]. The connection of this theory with the theory of algebras was given in Brauer [4], [5]. The $\mathfrak{p}$-indices were introduced in Hasse [16], [17], Brauer-Hasse-Noether [11]. See also the presentation of the theory of algebras in the books Deuring [15], Albert [1], Weyl [28], Jacobson [18], van der Waerden [26], Artin-Nesbitt-Thrall [2], and of the arithmetic parts of the theory in the first two of these books.

[8] See footnote 5.

[9] See for instance [13], § 6.

[10] Brauer [7], § 6. See also the arrangement of the proofs in Artin-Nesbitt-Thrall [2], Chapter IX, 8. An equivalent theorem is given in Nakayama [21], § 4.

with coefficients in the $\mathfrak{p}$-adic extension field $K(\mathfrak{p})$ of K with the following properties

(a) $\mathfrak{W}_\lambda(G)$ has $\mathfrak{p}$-integral coefficients for every G in $\mathfrak{G}$.

(b) If every coefficient of $\mathfrak{W}_\lambda(G)$ is replaced by its residue class (mod $\mathfrak{p}$), a modular representation $\mathfrak{W}^*_\lambda$ is obtained which is an indecomposable constituent $\mathfrak{U}_\lambda$ of the modular regular representation and which has the irreducible bottom constituent $\mathfrak{F}_\lambda$. Here, λ can have any of the values $1, 2, \cdots, l$.

Since $\mathfrak{F}_\lambda$ is absolutely irreducible, the representation $\mathfrak{U}_\lambda$ will remain indecomposable in any extension field Λ^* of K^*.[11] If K is replaced by an extension field Λ of finite degree, if $\mathfrak{p}$ is replaced by a prime ideal divisor $\mathfrak{q}$ of $\mathfrak{p}$, and if $K(\mathfrak{p})$ is replaced by the $\mathfrak{q}$-adic extension field $\Lambda(\mathfrak{q})$, the representation $\mathfrak{W}_\lambda$ will retain the properties (a) and (b). For instance, as in Section 1, we may take Λ as a field in which $\mathfrak{Z}_1, \mathfrak{Z}_2, \cdots, \mathfrak{Z}_k$ can be written. Then $\mathfrak{Z}_\kappa$ appears with the multiplicity $d_{\kappa\lambda}$ in $\mathfrak{W}_\lambda$,[12] and therefore,[13] the Schur index $\mu(\mathfrak{p})$ of $\mathfrak{Z}_\kappa$ with regard to $K(\mathfrak{p})$ is a divisor of $d_{\kappa\lambda}$. This index is termed the *$\mathfrak{p}$-index* of $\mathfrak{Z}_\kappa$ with regard to K.

We thus have

LEMMA 1. *If* K *is obtained from the field of rational numbers by adjoining the modular characters* $\phi_1, \phi_2, \cdots, \phi_l$ *and the character* ζ_κ *of the representation* $\mathfrak{Z}_\kappa$, *then the* $\mathfrak{p}$-*index* $\mu(\mathfrak{p})$ *of* $\mathfrak{Z}_\kappa$ *with regard to* K *divides all decomposition numbers* $d_{\kappa\lambda}$ *with the fixed first subscript* κ.

4. On the power of p dividing the $\mathfrak{p}$-index of $\mathfrak{Z}_\kappa$. Let p^a be the highest power of p dividing g, so that

$$g = p^a g', \qquad (g', p) = 1. \tag{4}$$

Every $\phi_\lambda(R)$ for p-regular R is a sum of g'-th roots of unity. Hence the field K_0 in (2) is contained in the field T of the g'-th roots of unity. Now (3) implies

$$K \subseteq T(\zeta_\kappa). \tag{5}$$

---

[11] As $\mathfrak{F}_\lambda$ is absolutely irreducible, the degree of the corresponding indecomposable constituent of the regular representation is equal to the multiplicity of $\mathfrak{F}_\lambda$ in the second regular representation, see [7], § 2 or [2]. This shows that $\mathfrak{U}_\lambda$ can not be decomposable in Λ^*.

[12] Brauer [7], equation (14). The proof of Lemma 1 can also be obtained directly from [7], equation (11), which is identical with Artin-Nesbitt-Thrall [2], equation (8.6) in Chapter IX.

[13] Schur [22], Theorem IV.

On the other hand, if ϵ denotes a primitive p^a-th root of unity, then $\Omega = \mathrm{T}(\epsilon)$ is the field of the g-th roots of unity, and we have

$$\mathrm{T} \subseteq \mathrm{T}(\zeta_\kappa) \subseteq \Omega = \mathrm{T}(\epsilon). \tag{6}$$

The degree r of $\mathrm{T}(\zeta_\kappa)$ with regard to T is equal to the number of p-conjugate characters to ζ_κ (that is, the number of distinct characters into which ζ_κ can be carried by an automorphism of Ω leaving the g'-th roots of unity fixed). If z_κ is the degree of $\mathfrak{Z}_\kappa$ then, as shown previously,[14]

$$r z_\kappa \not\equiv 0 \pmod{p^{a+1}}, \tag{7}$$

$$r z_\kappa \not\equiv 0 \pmod{p^a}, \quad \text{if} \quad z_\kappa \not\equiv 0 \pmod{p^a}. \tag{8}$$

If $z_\kappa \equiv 0 \pmod{p^a}$, a suitable $d_{\kappa\lambda}$ has the value 1,[15] and Lemma 1 shows that $\mu(\mathfrak{p}) \not\equiv 0 \pmod{p}$.

Assume now that $z_\kappa \not\equiv 0 \pmod{p^a}$. Then (8) can be applied. The $\mathfrak{p}$-index $\mu(\mathfrak{p})$, which is the Schur index of a representation $\mathfrak{Z}_\kappa$, divides the degree z_κ of this representation.[16] Hence

$$r\mu(\mathfrak{p}) \not\equiv 0 \pmod{p^a}. \tag{9}$$

The field Ω of the g-th roots of unity has the degree

$$[\Omega : \mathrm{T}] = p^{a-1}(p-1)$$

over the field T of the g'-th roots of unity, cf. (4). As the degree of $\mathrm{T}(\zeta_\kappa)$ over T was r, it follows that the degree of Ω over $\mathrm{T}(\zeta_\kappa)$ is given by

$$[\Omega : \mathrm{T}(\zeta_\kappa)] = p^{a-1}(p-1)/r.$$

Let $\mathfrak{P}$ be a prime ideal divisor of $\mathfrak{p}$ in Ω and let $\mathfrak{r}$ and $\mathfrak{r}_1$ be the prime ideals of T and $\mathrm{T}(\zeta_\kappa)$, respectively, which are divisible by $\mathfrak{P}$. Then $p \equiv 0 \pmod{\mathfrak{r}}$, and it is well known that

$$\mathfrak{r} = \mathfrak{P}^h \quad \text{with} \quad h = [\Omega : \mathrm{T}].$$

The relation (6) implies

$$\mathfrak{r}_1 = \mathfrak{P}^j \quad \text{with} \quad j = [\Omega : \mathrm{T}(\zeta_\kappa)],$$

and it now follows from (5) that if $\mathfrak{P}^e$ is the highest power of $\mathfrak{P}$ dividing the prime ideal $\mathfrak{p}$, the exponent e is divisible by $j = [\Omega : \mathrm{T}(\zeta_\kappa)] = p^{a-1}(p-1)/r$.

---

[14] Brauer [10], Theorem 2, p. 943.

[15] Brauer-Nesbitt [13], Theorem 1, p. 565.

[16] Schur [22], Theorem V.

The $\mathfrak{p}$-degree of Ω over K (that is the degree $n(\mathfrak{p})$ of the $\mathfrak{P}$-adic extension field $\Omega(\mathfrak{P})$ over $\mathrm{K}(\mathfrak{p})$),[17] is divisible by e. Hence $n(\mathfrak{p})$ is divisible by $p^{a-1}(p-1)/r$, and this gives

$$rn(\mathfrak{p}) \equiv 0 \pmod{p^{a-1}}. \tag{10}$$

On comparing (9) and (10), we obtain

LEMMA 2. *In the notation of Lemma* 1, *the highest power of p dividing $\mu(\mathfrak{p})$ also divides the $\mathfrak{p}$-degree $n(\mathfrak{p})$ of the field Ω of the g-th roots of unity over* K.

5. A lemma concerning the index of $\mathfrak{Z}_\kappa$ with regard to Ω. Choose a system of elements

$$P_0 = 1, P_1, P_2, \cdots, P_s\,^{18} \tag{11}$$

in $\mathfrak{G}$, such that (a) the order of every P_i is a power of p, (b) no two of these elements are conjugate in $\mathfrak{G}$, (c) every element of an order p^a is conjugate to one of the P_i. Set

$$\mathfrak{N}_i = \mathfrak{N}(P_i), \tag{12}$$

and determine a complete system of elements

$$V_1{}^i, V_2{}^i, \cdots \tag{13}$$

representing the different classes of p-regular elements in $\mathfrak{N}_i$. The elements $P_iV_j{}^i$ for all possible indices i, j form a system of representatives for the classes of conjugate elements in $\mathfrak{G}$.

The representation $\mathfrak{Z}_\kappa$ of $\mathfrak{G}$ induces a representation $\mathfrak{Z}_\kappa(\mathfrak{N}_i)$ of the subgroup $\mathfrak{N}_i$. If $\mathfrak{A}$ is an irreducible constituent, it represents the invariant element P_i of $\mathfrak{N}_i$ by a matrix of the form $\epsilon^\nu I$ where ϵ again denotes a primitive p^a-th root of unity. Hence

$$\mathfrak{A}(P_iV_j{}^i) = \mathfrak{A}(P_i)\mathfrak{A}(V_j{}^i) = \epsilon^\nu\mathfrak{A}(V_j{}^i)$$

which implies

$$\operatorname{tr}\mathfrak{A}(P_iV_j{}^i) = \epsilon^\nu \operatorname{tr}\mathfrak{A}(V_j{}^i). \tag{14}$$

The character $\operatorname{tr}\mathfrak{A}$ can be expressed by the irreducible modular characters $\phi_1{}^i, \phi_2{}^i, \cdots$ of $\mathfrak{N}_i$. We thus see that the left side in (14) is a linear combination of $\phi_1{}^i(V_j{}^i), \phi_2{}^i(V_j{}^i), \cdots$ with coefficients which are independent of j. These coefficients are non-negative rational integers multiplied by ϵ^{ν}.

---

[17] See, for instance, Hasse [17], (3.3).

[18] For the following, cf. Brauer [9], Section 1.

Adding the formulae (14) for all irreducible constituents of $\mathfrak{Z}_\kappa(\mathfrak{N}_i)$, each taken with the correct multiplicity, we finally obtain formulae

$$\zeta_\kappa(P_i V_j{}^i) = \sum_\rho d_{\kappa\rho}{}^i \phi_\rho{}^i(V_j{}^i) \tag{15}$$

where the $d_{\kappa\rho}{}^i$ are algebraic integers, the *generalized decomposition numbers* of $\mathfrak{G}$.

We now prove

LEMMA 3. *Assume that all irreducible representations of $\mathfrak{N}_i = \mathfrak{N}(P_i)$ (for a fixed i) can be written in a field Σ. The Schur index m of $\mathfrak{Z}_\kappa$ with regard to Σ divides $d_{\kappa\rho}{}^i$ for all values of ρ.*

Proof. If $\mathfrak{N}_i = \mathfrak{G}$, the assumption implies that $m = 1$, and the statement is trivial. If $\mathfrak{N}_i \subset \mathfrak{G}$, then it follows from a theorem of I. Schur,[19] that m divides the multiplicity of every constituent $\mathfrak{A}$ of $\mathfrak{Z}_\kappa(\mathfrak{N}_i)$. In deriving (15) from (14) we then have to multiply each equation (14) by a multiple of m before adding. This gives the statement of Lemma 3.

6. Proof of the theorem. Because of the complete reducibility of the representations of $\mathfrak{G}$, it is sufficient to prove the theorem formulated in the introduction for absolutely irreducible representations $\mathfrak{L}$. Assume that the theorem is false. Take a group $\mathfrak{G}$ of minimal order g for which there exists an absolutely irreducible representation $\mathfrak{L}$ which cannot be written in the field Ω of the g-th roots of unity. In other words, the Schur index m of $\mathfrak{L}$ with regard to Ω is different from 1. The representation $\mathfrak{L}$ is similar to one of the $\mathfrak{Z}_\kappa$; we may assume $\mathfrak{L} = \mathfrak{Z}_\kappa$ for a certain κ. This means[20] that there exists a prime ideal $\mathfrak{P}$ of Ω, finite or infinite, such that the $\mathfrak{P}$-index $\mathrm{M}(\mathfrak{P})$ of $\mathfrak{Z}_\kappa$ with regard to Ω is different from 1,

$$\mathrm{M}(\mathfrak{P}) > 1. \tag{16}$$

Since g must be larger than 2, the field Ω has no real conjugate. Then $\mathfrak{P}$ in (16) can not be infinite.

Assume that $\mathfrak{P}$ is finite. Let K have the same significance as in Lemma 1. Then $\mathrm{K} \subseteq \Omega$. Let $\mathfrak{p}$ be the prime ideal of K divisible by $\mathfrak{P}$, and again let p be the rational prime number divisible by $\mathfrak{p}$.

If for an element P_i in (11) we have $\mathfrak{N}_i \subset \mathfrak{G}$, the group $\mathfrak{N}_i$ has an order which is a proper divisor of $\mathfrak{G}$. The theorem is correct for $\mathfrak{N}_i$, and all representations of $\mathfrak{N}_i$ can be written in Ω. Lemma 3 yields

---

[19] Schur [22], Theorem IX a. See also Brauer [6], Theorem IV.

[20] Brauer-Hasse-Noether [11], Theorem I.

$$(17)\qquad d_{\kappa\lambda}{}^i \equiv 0 \pmod{m}$$

for all λ.

On the other hand, if $\mathfrak{N}(P_i) = \mathfrak{G}$ (as, for instance, for $i = 0$, $P_0 = 1$), the constituent $\mathfrak{A}$ of $\mathfrak{Z}_\kappa(\mathfrak{N}_i) = \mathfrak{Z}_\kappa(\mathfrak{G})$ in Section **5** coincides with $\mathfrak{Z}_\kappa$. Hence (14) reads

$$\zeta_\kappa(P_i V_j{}^i) = \epsilon^\nu \zeta_\kappa(V_j{}^i).$$

The equation (1) now yields

$$(18)\qquad \zeta_\kappa(P_i V_j{}^i) = \epsilon^\nu \sum_\rho d_{\kappa\rho}\phi_\rho(V_j{}^i).$$

The irreducible modular characters $\phi_1{}^i, \phi_2{}^i, \cdots$ of $\mathfrak{N}_i = \mathfrak{G}$ are the characters $\phi_1, \phi_2, \cdots$; we may assume $\phi_\rho{}^i = \phi_\rho$. On comparing (15) and (18), we have $d_{\kappa\rho}{}^i = \epsilon^\nu d_{\kappa\rho}$. Now Lemma 1 shows that

$$(19)\qquad d_{\kappa\rho}{}^i \equiv 0 \pmod{\mu(\mathfrak{p})}.$$

In any case, every $d_{\kappa\rho}{}^i$ with the fixed first subscript κ is divisible by the greatest common divisor of $\mu(\mathfrak{p})$ and m.

The orthogonality relation for group characters gives

$$(20)\qquad (1/g)\sum \zeta_\kappa(G)\overline{\zeta_\kappa(G)} = 1$$

where G ranges over all elements of $\mathfrak{G}$. The elements $P_i V_j{}^i$ form a complete system of representatives for the classes of conjugate elements in $\mathfrak{G}$. The class of $P_i V_j{}^i$ contains $g/n(P_i V_j{}^i)$ elements. Now (20) can be written in the form

$$\sum_i \sum_j \zeta_\kappa(P_i V_j{}^i)\overline{\zeta_\kappa(P_i V_j{}^i)}/n(P_i V_j{}^i) = 1.$$

Combining this with (15), we have

$$(21)\qquad \sum_i \sum_j \sum_\rho \sum_\sigma d_{\kappa\rho}{}^i \overline{d_{\kappa\sigma}{}^i}\phi_\rho{}^i(V_j{}^i)\overline{\phi_\sigma{}^i(V_j{}^i)}/n(P_i V_j{}^i) = 1.$$

In the group $\mathfrak{N}_i$ of order $n(P_i)$, the number of elements in the class of $V_j{}^i$ is $n(P_i)/n(P_i V_j{}^i)$ since every element of $\mathfrak{G}$ commuting with $P_i V_j{}^i$ automatically lies in $\mathfrak{N}_i = \mathfrak{N}(P_i)$. Hence the orthogonality relation for the modular group characters[21] of $\mathfrak{N}_i$ reads

$$(22)\qquad \sum_j \phi_\rho{}^i(V_j{}^i)\overline{\phi_\sigma{}^i(V_j{}^i)}/n(P_i V_j{}^i) = \gamma_{\rho\sigma}{}^i$$

---

[21] Brauer-Nesbitt **[13]**, (21).

where the matrix $(\gamma_{\rho\sigma}^i)$ is the inverse of the matrix $C^{(i)}$ of the Cartan-invariants of $\mathfrak{N}_i$.

On account of (22), the equation (21) can be written in the form

$$\sum_i \sum_\rho \sum_\sigma d_{\kappa\rho}^i \overline{d_{\kappa\sigma}^i} \gamma_{\rho\sigma}^i = 1.$$

For a suitable rational integer t the products $p^t\gamma_{\rho\sigma}^i$ are rational integers.[22] We write the last equation in the form

$$\sum_i \sum_\rho \sum_\sigma d_{\kappa\rho}^i \overline{d_{\kappa\sigma}^i} (p^t\gamma_{\rho\sigma}^i) = p^t. \tag{23}$$

Since every $d_{\kappa\rho}^i$ was divisible by the greatest common divisor of $\mu(\mathfrak{p})$ and m, it follows, from (23), that the square of this greatest common divisor divides p^t. In any case, this greatest common divisor $(m, \mu(\mathfrak{p}))$ is a power of p.

The $\mathfrak{P}$-index $\mathrm{M}(\mathfrak{P})$ of $\mathfrak{Z}_\kappa$ with regard to Ω divides both the $\mathfrak{p}$-index of $\mathfrak{Z}_\kappa$ with regard to K and the index m of $\mathfrak{Z}_\kappa$ with regard to Ω. Hence

$$\mathrm{M}(\mathfrak{P}) = p^b \tag{24}$$

for a certain integral rational exponent $b \geqq 0$.

The absolutely irreducible representation $\mathfrak{Z}_\kappa$ has the index $\mu(\mathfrak{p})$ with respect to the field $\mathrm{K}(\mathfrak{p})$ which contains its character ζ_κ. Then $\mathfrak{Z}_\kappa$ determines a normal simple algebra A of index $\mu(\mathfrak{p})$ over $\mathrm{K}(\mathfrak{p})$ which consists of all linear combinations of elements of $\mathfrak{Z}_\kappa$ with coefficients in $\mathrm{K}(\mathfrak{p})$. The Hasse invariant[23] of A is a rational number $\rho = q/\mu(\mathfrak{p})$ with the denominator $\mu(\mathfrak{p})$, $(q, \mu(\mathfrak{p})) = 1$. If $\mathrm{K}(\mathfrak{p})$ is replaced by the extension field $\Omega(\mathfrak{P})$, then ρ is multiplied[24] by the relative degree $n(\mathfrak{p}) = [\Omega(\mathfrak{P}) : \mathrm{K}(\mathfrak{p})]$. Therefore, the new index $M(\mathfrak{P})$ is the denominator of $\rho\, n(\mathfrak{p}) = qn(\mathfrak{p})/\mu(\mathfrak{p})$ after numerator and denominator have been freed of their common divisor. Hence

$$\mathrm{M}(\mathfrak{P}) = \mu(\mathfrak{p})/(\mu(\mathfrak{p}), n(\mathfrak{p})). \tag{25}$$

Combination of (25) with Lemma 2 shows that $\mathrm{M}(\mathfrak{P})$ is not divisible by p, and (24) then implies that $\mathrm{M}(\mathfrak{P}) = 1$, in contradiction to (16). Hence, the assumption $m \neq 1$ leads to a contradiction, and the theorem is proved.

---

[22] The coefficients of $C^{(i)}$ are rational integers; the determinant of $C^{(i)}$ is a power of p, Brauer [8]. This shows already that $\gamma_{\rho\sigma}^i$ is a rational number whose denominator is a power of p. As can be seen easily, t can be chosen equal to a, but this is not needed in the following.

[23] Hasse [17], (5.1). See also Deuring [15], Chapter VII, §§ 2, 4.

[24] Köthe [19], Hasse [17], (5.2), also Deuring [15], Chapter VII, § 2.

7. The case of reducible representations. The first remark in Section 6 shows that the theorem also holds for reducible ordinary representations $\mathfrak{R}$. However, this is no longer true in the case of modular representations. For instance, if Π_0 denotes the Galois field with two elements and if t is a transcendental element with regard to Π_0 the four matrices

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},\quad \begin{pmatrix}1&0&0\\0&1&0\\1&0&1\end{pmatrix},\quad \begin{pmatrix}1&0&0\\0&1&0\\t&0&1\end{pmatrix},\quad \begin{pmatrix}1&0&0\\0&1&0\\t+1&0&1\end{pmatrix}$$

form a group of order four which can not be written in Π_0 as can be shown without difficulty.

UNIVERSITY OF TORONTO.

BIBLIOGRAPHY.

[1]. A. A. Albert, *Structure of Algebras*, American Mathematical Society Colloquium Publications, vol. 24 (1939).

[2]. E. Artin, C. J. Nesbitt, and R. M. Thrall, "Rings with minimum condition," *University of Michigan Publications in Mathematics*, No. 1, Ann Arbor, 1944.

[3]. R. Brauer, "Ueber Zusammenhänge zwischen arithmetischen und invariantentheoretischen Eigenschaften von Gruppen linearer Substitutionen," *Sitz. Ber. Preuss. Akad. Wiss.* (1926), pp. 410-416.

[4]. R. Brauer, "Untersuchungen über die arithmetischen Eigenschaften von Gruppen linearer Substitutionen, I," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 677-696.

[5]. R. Brauer, "Ueber Systeme hyperkomplexer Zahlen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 79-107.

[6]. R. Brauer, "Untersuchungen über die arithmetischen Eigenschaften von Gruppen linearer Substitutionen, II," *Mathematische Zeitschrift*, vol. 31 (1930), pp. 737-747.

[7]. R. Brauer, "On modular and p-adic representations of algebras," *Proceedings of the National Academy of Sciences*, vol. 25 (1939), pp. 252-258.

[8]. R. Brauer, "On the Cartan invariants of groups of finite order," *Annals of Mathematics*, vol. 42 (1941), pp. 53-61.

[9]. R. Brauer, "On the connection between the ordinary and the modular characters of groups of finite order," *Annals of Mathematics*, vol. 42 (1941), pp. 926-935.

[10]. R. Brauer, "Investigations on group characters," *Annals of Mathematics*, vol. 42 (1941), pp. 936-958.

[11]. R. Brauer, H. Hasse, and E. Noether, "Beweis eines Hauptsatzes in der Theorie der Algebren," *Jour. reine angew. Math.*, vol. 167 (1932), pp. 399-404.

[12]. R. Brauer and C. J. Nesbitt, *On the Modular Representations of Groups of Finite Order*, Toronto Studies, 1937.

[13]. R. Brauer and C. J. Nesbitt, "On the modular characters of groups," *Annals of Mathematics*, vol. 42 (1941), pp. 556-590.
[14]. W. Burnside, "On the complete reduction of any transitive permutation group and on the arithmetic nature of the coefficients in its irreducible components," *Proceedings of the London Mathematical Society* (2), vol. 3 (1905), pp. 239-252.
[15]. M. Deuring, "Algebren," *Ergebnisse der Math.*, vol. 4 (Berlin, 1935).
[16]. H. Hasse, "Theory of cyclic algebras over an algebraic number field," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 171-214, and pp. 727-730.
[17]. H. Hasse, "Die Struktur der R. Brauerschen Algebrenklassengruppe über einem algebraischen Zahlkörper," *Mathematische Annalen*, vol. 107 (1933), pp. 731-760.
[18]. N. Jacobson, "The Theory of Rings," *Mathematical Surveys*, vol. II, 1943.
[19]. G. Käthe, "Erweiterung des Zentrums einfacher Algebren," *Mathematische Annalen*, vol. 107. (1933), pp. 761-766.
[20]. H. Maschke, "Ueber den arithmetischen Charakter der Coefficienten der Substitutionen endlicher linearer Substitutionsgruppen," *Mathematische Annalen*, vol. 50 (1898), pp. 482-498.
[21]. T. Nakayama, "Some studies on regular representations, induced representations, and modular representations," *Annals of Mathematics*, vol. 39 (1938), pp. 361-369.
[22]. I. Schur, "Arithmetische Untersuchungen über endliche Gruppen linearer Substitutionen," *Sitz. Ber. Preuss. Akad. Wiss.*, 1906, pp. 164-184.
[23]. I. Schur, "Beiträge zur Theorie der Gruppen linearer homogener Substitutionen," *Transactions of the American Mathematical Society*, vol. 15 (1909), pp. 159-175.
[24]. I. Schur, "Einige Bemerkungen zu der vorstehenden Arbeit des Herrn A. Speiser," *Mathematische Zeitschrift*, vol. 5 (1919), pp. 6-10.
[25]. A. Speiser, "Zahlentheoretische Sätze aus der Gruppentheorie," *Mathematische Zeitschrift*, vol. 5 (1919), pp. 1-6.
[26]. B. L. van der Waerden, *Moderne Algebra*, vol. II, 2nd edition, Berlin, 1940.
[27]. J. H. M. Wedderburn, "A theorem on finite algebras," *Transactions of the American Mathematical Society*, vol. 6 (1905), pp. 349-352.
[28]. H. Weyl, "The Classical Groups," Princeton, 1939.

EIN SATZ UEBER QUADRATISCHE FORMEN MIT KOMPLEXEN KOEFFIZIENTEN.*

By Issai Schur.

Im folgenden werde ich den Satz beweisen:

I. *Jede quadratische Form mit reellen oder komplexen Koeffizienten* $f(x) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa x_\lambda$ $(c_{\kappa\lambda} = c_{\lambda\kappa})$ *laesst sich in der Gestalt*

$$f(x) = \sum_{\nu=1}^{n} \omega_\nu \phi_\nu(x)^2, \qquad \phi_\nu(x) = \sum_{\alpha=1}^{n} d_{\nu\alpha} x_\alpha \tag{1}$$

darstellen, wobei $\omega_1, \cdots, \omega_n$ *nicht-negative reelle Zahlen sind und*

$$| \phi_1 |^2 + | \phi_2 |^2 + \cdots + | \phi_n |^2 = | x_1 |^2 + | x_2 |^2 + \cdots + | x_n |^2 \tag{2}$$

oder, was dasselbe ist,

$$\sum_{\alpha=1}^{n} d_{\kappa\alpha} \bar{d}_{\lambda\alpha} = e_{\kappa\lambda}, \quad \sum_{\beta=1}^{n} d_{\beta\kappa} \bar{d}_{\beta\lambda} = e_{\kappa\lambda} \quad (e_{\kappa\kappa} = 1,\ e_{\kappa\lambda} = 0 \text{ für } \kappa \neq \lambda). \tag{3}$$

Mit anderen Worten:

I. *Jede symmetrische Matrix* $C = (c_{\kappa\lambda})$ *laesst sich mit Hilfe einer unitären Substitution auf die Form*

$$UCU' = \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \omega_n \end{pmatrix} \tag{4}$$

bringen, wo $\omega_1, \cdots, \omega_n$ *nichtnegative reelle Zahlen sind.*

Dies bedeutet genauer: Jede symmetrische Bilinearform $f(x, y) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa y_\lambda$ laesst sich in der Form

$$f(x, y) = \sum_{\nu=1}^{n} \omega_\nu \phi_\nu(x) \phi_\nu(y) \tag{5}$$

darstellen, wobei die $\phi_\nu(x) = \sum_{\alpha=1}^{n} d_{\nu\alpha} x_\alpha$ den Bedingungen (2) bezw. (3) genügen.

---

* Received November 8, 1943. This paper has been edited, after the author's death, by Professors Fekete and M. Schiffer of the University of Jerusalem.

1. **Bezeichnungen.** Jeder Matrix n-ten Grades $C = (c_{\kappa\lambda})$ $(\kappa, \lambda = 1, \cdots, n)$ entspricht einerseits die lineare Substitution (C): $x_\kappa' = \sum_{\lambda=1}^{n} c_{\kappa\lambda} x_\lambda$ $(\kappa = 1, 2, \cdots, n)$ mit unabhängigen Veränderlichen $x_1, \cdots, x_n$, andererseits die Bilinearform

$$C(x, y) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa y_\lambda. \tag{6}$$

Für (C) führen wir die Bezeichnung ein: Ist $z = \sum_{\nu=1}^{n} \xi_\nu x_\nu$ eine beliebige Linearform, so bedeutet $z' = z^C$ die Linearform $z' = z^C = \sum_{\nu=1}^{n} \xi_\nu (\sum_{\alpha=1}^{n} c_{\nu\alpha} x_\alpha)$ $= \sum_{\alpha=1}^{n} x_\alpha (\sum_{\nu=1}^{n} c_{\nu\alpha} \xi_\nu)$. Wird (identisch in den x_α)

$$z' = \omega z \quad (\omega \text{ konstant}) \tag{7}$$

und $z \not\equiv 0$, so heisst (bekanntlich) z eine Invariante der Substitution (C). Derartige Invarianten gibt es bekanntlich dann und nur dann, wenn für ω eine charakteristische Wurzel der Gleichung

$$| C - xE | = | c_{\kappa\lambda} - x e_{\kappa\lambda} | = 0$$

gewählt wird.

Zu jeder nicht identisch verschwindenden Linearform z lassen sich n linear unabhängige Linearformen

$$z_\kappa = \sum_{\lambda=1}^{n} s_{\kappa\lambda} x_\lambda, \quad (\kappa = 1, \cdots, n), \qquad | s_{\kappa\lambda} | \neq 0, \quad z_1 = z$$

wählen, und zwar auch solche, die eine unitäre (bezw. reell orthogonale) Substitution bestimmen, wenn z "normiert" ist, *d. h.* $\sum_{\nu=1}^{n} | \xi_\nu |^2 = 1$ ist, bezw. wenn die ξ_ν reell sind und $\sum_{\nu=1}^{n} \xi_\nu^2 = 1$ wird. Bildet man für jedes κ die transformierte Linearform $z'_\kappa = z_\kappa^C = \sum_{\lambda=1}^{n} s_{\kappa\lambda} (\sum_{\alpha=1}^{n} c_{\lambda\alpha} x_\alpha)$ und drückt man die z'_κ durch die z_κ aus, so erhält man in (C_1): $z'_\kappa = \sum_{\lambda=1}^{n} c'_{\kappa\lambda} z_\lambda$ $(\kappa = 1, \cdots, n)$ eine zu (C) ähnliche Substitution (C_1), mit der Matrix

$$C_1 = SCS^{-1} \qquad (S = (s_{\kappa\lambda})). \tag{8}$$

Für die Bilinearform $C(x, y)$ gilt ferner folgendes: Ist $S = (s_{\kappa\lambda})$ irgend eine Matrix (Substition) und ersetzt man simultan ("kogredient") die x_κ, y_λ durch $x_\kappa = \sum_{\mu=1}^{n} s_{\mu\kappa} u_\mu$, $y_\lambda = \sum_{\nu=1}^{n} s_{\nu\lambda} v_\nu$, so wird

$$C(x, y) = \sum_{\kappa,\lambda,\mu,\nu} c_{\kappa\lambda} s_{\mu\kappa} s_{\nu\lambda} u_\mu v_\nu = \sum_{\mu,\nu} c_{\mu\nu}'' u_\mu v_\nu$$

wo

$$(9) \qquad C_2 = (c_{\mu\nu}'') = SCS'.$$

wird. Hierbei bedeutet S' die zu S "konjugierte," die "transponierte" Matrix. (N. B. Nur dann, wenn S orthogonal ist, wird dies eine Aehnlichkeitstransformation, d. h. $S' = S^{-1}$, nicht dagegen, wenn S nur als unitär, aber komplex bekannt ist, denn alsdann ist nur $S' = \bar{S}^{-1}$.) Im folgenden bezeichne ich in bekannter Weise, wenn C_1, C_2, C_3, C_4 vier Matrizen n-ten Grades $C_a = (c_{\kappa\lambda}{}^{(a)})$ sind, mit $\begin{pmatrix} C_1 & C_2 \\ C_3 & C_4 \end{pmatrix}$ die Matrix des Grades $2n$, die ausführlicher geschrieben, die Gestalt

$$M = \begin{pmatrix} c_{11}{}^{(1)} & \cdots & c_{1n}{}^{(1)} & c_{11}{}^{(2)} & \cdots & c_{1n}{}^{(2)} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ c_{n1}{}^{(1)} & \cdots & c_{nn}{}^{(1)} & c_{n1}{}^{(2)} & \cdots & c_{nn}{}^{(2)} \\ c_{11}{}^{(3)} & \cdots & c_{1n}{}^{(3)} & c_{11}{}^{(4)} & \cdots & c_{1n}{}^{(4)} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ c_{n1}{}^{(3)} & \cdots & c_{nn}{}^{(3)} & c_{n1}{}^{(4)} & \cdots & c_{nn}{}^{(4)} \end{pmatrix}$$

hat. Mit diesen "zusammengesetzten" Bildungen kann man im Kalkül additiv und multiplikativ bekanntlich ebenso operieren wie mit Zahlenmatrizen, nur dass man bei der Multiplikation auf die Reihenfolge der Faktoren achtgeben muss. Man beachte insbesondere, dass die Transponierte von M gleich $M' = \begin{pmatrix} C'_1 & C'_3 \\ C'_2 & C'_4 \end{pmatrix}$ zu setzen ist.

2. Der Stickelbergersche Beweis für das Hauptachsentheorem. Um das Folgende verständlicher zu machen, gebe ich hier den schönen, klassisch gewordenen, Stickelbergerschen Beweis für das Hauptachsentheorem wieder. Ist $C = (c_{\kappa\lambda})$ eine *reelle* symmetrische Matrix und ω eine charakteristische Wurzel von C, so bestimme man eine zugehörige invariante Linearform $z = \sum_{\nu=1}^{n} \xi_\nu x_\nu$ mit $z^C = \omega z$, was auf

$$(10) \qquad \sum_{\nu=1}^{n} c_{\nu a} \xi_\nu = \omega \xi_a$$

führt. Die $\xi_1, \cdots, \xi_n$ brauchen zunächst noch nicht reell zu sein, man darf aber den noch unbestimmt bleibenden Proportionalitätsfaktor so wählen, dass $|\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2=1$ wird. Aus (10) folgt aber (Cauchy!) wegen $c_{\nu a}=c_{a\nu}=\bar{c}_{\nu a}$

$$\omega=\omega\sum_{a=1}^{n}|\xi_a|^2=\sum_{\nu,a}c_{\nu a}\xi_\nu\bar{\xi}_a=\sum_{\nu,a}c_{a\nu}\xi_\nu\bar{\xi}_a$$

und

$$\bar{\omega}=\sum_{\nu,a}c_{\nu a}\bar{\xi}_\nu\xi_a=\sum_{\nu,a}c_{a\nu}\bar{\xi}_a\xi_\nu=\omega;$$

also ist ω reell, und folglich darf man $\xi_1,\cdots,\xi_n$ auch reell wählen. Wählt man, was jedenfalls möglich ist, n Linearformen $z_\kappa=\sum_{\lambda=1}^{n}s_{\kappa\lambda}x_\lambda$, die eine reelle orthogonale Transformation bestimmen, so dass $z_1=z$ wird, so führt die Einführung von $z_1, z_2,\cdots,z_n$ als neue Veränderliche auf eine Aehnlichkeitstransformation (vgl. 1) $SCS^{-1}=(c'_{\kappa\lambda})$, worin wegen $z'=\omega z$, $z=z_1$ in der ersten Zeile $c'_{11}=\omega_1$, $c'_{12}=\cdots=c'_{1n}=0$ wird. Da aber $S^{-1}=S'$ wird, erhalten wir in SCS^{-1} auch SCS', d. h. von neuem Symmetrie, also $c'_{\kappa\lambda}=c'_{\lambda\kappa}$;

insbesondere wird $SCS^{-1}=\begin{pmatrix}\omega & 0 & \cdot & \cdot & \cdot & 0\\ 0 & d_{22} & \cdot & \cdot & \cdot & d_{2n}\\ \cdot & \cdot & & & & \cdot\\ \cdot & \cdot & & & & \cdot\\ \cdot & \cdot & \cdot & & & \cdot\\ 0 & d_{n2} & \cdot & \cdot & \cdot & d_{nn}\end{pmatrix}$. Die symmetrische reelle Matrix $(d_{\nu\mu})$ kann man nun weiterhin ebenso behandeln usw.

3. Der Haupthilfssatz. Ist $U=(u_{\kappa\lambda})$ eine unitäre Substitution n-ten Grades, so setze man

$$u_{\kappa\lambda}=p_{\kappa\lambda}+iq_{\kappa\lambda},\qquad (p_{\kappa\lambda})=P,\quad (q_{\kappa\lambda})=Q.$$

Jede symmetrische Matrix $C=(c_{\kappa\lambda})$ mit $c_{\kappa\lambda}=a_{\kappa\lambda}+ib_{\kappa\lambda}$, $(a_{\kappa\lambda})=A$, $(b_{\kappa\lambda})=B$ liefert dann bei der Transformation UCU' eine neue Zerlegung in Real- und Imaginärteil $UCU'=A^*+iB^*$. Hierbei wird

$$UCU'=(P+iQ)(A+iB)(P'+iQ')=A^*+iB^*,$$

woraus folgt

$$\begin{aligned}A^*&=PAP'-QBP'-PBQ'-QAQ'\\ B^*&=PAQ'-QBQ'+PBP'+QAP'.\end{aligned}\tag{11}$$

Operiert man ferner bei der unitären Substitution U mit komplexen Veränderlichen $x_\nu=u_\nu+iv_\nu$, so liefert

$$x'_\kappa=u'_\kappa+iv'_\kappa=\sum_{\lambda=1}^{n}u_{\kappa\lambda}x_\lambda=\sum_{\lambda=1}^{n}(p_{\kappa\lambda}+iq_{\kappa\lambda})(u_\lambda+iv_\lambda)$$

die Formeln

$$(12)\qquad u'_\kappa = \sum_{\lambda=1}^{n} (p_{\kappa\lambda}u_\lambda - q_{\kappa\lambda}v_\lambda), \qquad v'_\kappa = \sum_{\lambda=1}^{n} (q_{\kappa\lambda}u_\lambda + p_{\kappa\lambda}v_\lambda).$$

Hier handelt es sich um eine lineare Transformation in den $2n$ reellen Veränderlichen, $u_1, \cdots, u_n, v_1, \cdots, v_n$ die wegen $\sum\limits_{\kappa=1}^{n} |x'_\kappa|^2 = \sum\limits_{\kappa=1}^{n} |x_\kappa|^2$ auf

$$u'^2_1 + \cdots + u'^2_n + v'^2_1 + \cdots + v'^2_n = u_1^2 + \cdots + u_n^2 + v_1^2 + \cdots + v_n^2,$$

also auf eine relle orthogonale Transformation führt. *Hält man an der Reihenfolge* $u_1, \cdots, u_n, v_1, \cdots, v_n$ *fest,* so erscheint als Matrix der Transformation die Matrix

$$(12')\qquad D = \begin{pmatrix} P & -Q \\ Q & P \end{pmatrix}.$$

Man bilde nun mit Hilfe der Matrix $C = A + iB$ die "zugeordnete" *reelle symmetrische Matrix des Grades* $2n$ $R = \begin{pmatrix} A & B \\ B & -A \end{pmatrix} = R(C) = R(A + iB)$. Es gilt dann der für uns ausschlaggebende Hilfssatz:

II. *Ist* U *eine unitäre Transformation* $U = P + iQ$ *in den* n *Veränderlichen* $x_1, \cdots, x_n$, *und bildet man neben* $C = A + iB$ *noch* $UCU' = A^* + iB^*$, *so hat die Bildung* $R(C)$ *die Grundeigenschaft*:

$$(13)\qquad DR(C)D' = R(UCU') = \begin{pmatrix} A^* & B^* \\ B^* & -A^* \end{pmatrix}.$$

Der Beweis beruht auf einer einfachen Rechnung: Es wird

$$DR(C)D' = \begin{pmatrix} P & -Q \\ Q & P \end{pmatrix} \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \begin{pmatrix} P' & Q' \\ -Q' & P' \end{pmatrix} = \begin{pmatrix} PA - QB, & PB + QA \\ QA + PB, & QB - PA \end{pmatrix} \begin{pmatrix} P' & Q' \\ -Q' & P' \end{pmatrix}$$

$$= \begin{pmatrix} PAP' - QBP' - PBQ' - QAQ', & PAQ' - QBQ' + PBP' + QAP' \\ QAP' + PBP' - QBQ' + PAQ', & QAQ' + PBQ' + QBP' - PAP' \end{pmatrix}.$$

Ein Vergleich mit den Formeln (11) liefert unseren Hilfssatz. Man beachte noch folgendes: Die Schreibweise $R = \begin{pmatrix} A & B \\ B & -A \end{pmatrix}$ ist an die Reihenfolge $u_1, \cdots, u_n,\ v_1 \cdots v_n$ geknüpft. Zieht man die natürliche Reihenfolge $u_1, v_1, \cdots, u_n, v_n$ vor, so erhält die reelle symmetrische Matrix R die Koeffizientenmatrix $R_1 = \begin{pmatrix} a_{11} & b_{11} & a_{12} & b_{12} \cdots \\ b_{11} & -a_{11} & b_{12} & -a_{12} \cdots \\ \cdot\ \cdot\ \cdot\ \cdot & & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdots \end{pmatrix}$. Hierbei geht R_1 aus R durch eine Aehnlichkeitstransformation hervor. Man beachte noch folgendes: Die charakteristischen Wurzeln einer Matrix der Form R haben (auch für *beliebige* zwei Teilmatrizen A und B) die Gestalt $\omega_1, \omega_2, \cdots, \omega_n, -\omega_1, -\omega_2, \cdots, -\omega_n$. Dies folgt am einfachsten aus der Formel

$$\begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix} \begin{pmatrix} A & B \\ B & -A \end{pmatrix} = \begin{pmatrix} -A & -B \\ -B & A \end{pmatrix} \begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix},$$

woraus hervorgeht, dass R und $-R$ einander aehnliche Matrizen sind, dass also die charakteristische Determinante von R eine gerade Funktion der Unbekannten ist. Hieraus folgt, dass R jedenfalls im symmetrischen Fall auch nichtnegative charakteristische Wurzeln besitzt.

4. **Beweis des Satzes I.** Für unsere symmetrische Matrix $C = (c_{\kappa\lambda})$, $c_{\kappa\lambda} = a_{\kappa\lambda} + ib_{\kappa\lambda}$ suchen wir diejenigen nicht identisch verschwindenden Linearformen $\psi = \sum_{\nu=1}^{n} \xi_\nu x_\nu$, für die die transformierte Funktion $\psi' = \sum_{\nu=1}^{n} \xi_\nu \sum_{a=1}^{n} c_{\nu a} x_a$ sich nur um einen konstanten Faktor $\bar{\omega}$ von $\bar{\psi} = \sum_{\nu=1}^{n} \bar{\xi}_\nu x_\nu$ unterscheidet, d. h. es soll $\psi' = \psi^C = \bar{\omega}\bar{\psi}$ sein. Hier kann nur der absolute Betrag von $\bar{\omega}$ bestimmt sein. Denn existiert eine solche Funktion und setzt man für alle $\xi_\nu = e^{i\phi}\xi_\nu^{(1)}$, so erhält man $e^{i\phi}(\sum_{\nu=1}^{n} \xi_\nu^{(1)} x_\nu)' = \bar{\omega} e^{-i\phi} \sum_{\nu=1}^{n} \overline{\xi_\nu^{(1)}} x_\nu$. Ist aber $|\bar{\omega}| = \omega$, $\bar{\omega} = \omega e^{i\alpha}$, so setze man $\phi = \alpha/2$. An Stelle von $\bar{\omega}$ tritt dann in $\psi' = \bar{\omega}\psi$ die reelle nichtnegative Grösse ω. Man betrachte also von vornherein nur die "Halbinvarianten" $\psi = \sum_{\nu=1}^{n} \xi_\nu x_\nu$, für die bei reellem nicht negativem ω, $\psi' = \omega\psi$ oder, was dasselbe ist,

$$\sum_{\nu=1}^{n} c_{\nu a}\xi_\nu = \omega\bar{\xi}_a \tag{14}$$

wird. Setzt man $\xi_\nu = \zeta_\nu - i\eta_\nu$, so erhält man

$$\sum_{\nu=1}^{n} (a_{\nu a} + ib_{\nu a})(\zeta_\nu - i\eta_\nu) = \omega(\zeta_a + i\eta_a)$$

also

$$\sum_{\nu=1}^{n} (a_{\nu a}\zeta_\nu + b_{\nu a}\eta_\nu) = \omega\zeta_a; \qquad \sum_{\nu=1}^{n} (b_{\nu a}\zeta_\nu - a_{\nu a}\eta_\nu) = \omega\eta_a. \tag{15}$$

Dies besagt aber nur, dass im symmetrischen Fall C Halbinvarianten mit reellem nichtnegativem ω bezitzt; es gehört nämlich zu jeder solchen „Wurzel" der zugeordneten Matrix $R = R(C)$ eine solche Halbinvariante von C. Genauer: Jede Halbinvariante $\psi = \sum_{\nu=1}^{n} (\zeta_\nu - i\eta_\nu)x_\nu$ von C bestimmt in $\Psi(u, v) = \sum_{\nu=1}^{n} \zeta_\nu u_\nu$

$+\sum_{\nu=1}^{n}\eta_\nu v_\nu$ eine zum gleichen Faktor gehörende Invariante der reellen Matrix $R=R(C)$. Um nun den Satz I zu beweisen, schliesse man so: Man bestimme eine Wurzel $\omega \geqq 0$ von $R=R(C)$ und eine zugehörige Halbinvariante $\psi=\sum_{\nu=1}^{n}\xi_\nu x_\nu$ von C. Diese ξ_ν darf man als normiert ansehen, da man sonst die ξ_ν durch $\dfrac{\xi_\nu}{\sqrt{|\xi_1|^2+\cdots+|\xi_n|^2}}$ ersetzen kann, was die Forderung an ψ nicht beeinflusst. Die Linearform ψ ergänze man zu einem System $\psi=\psi_1, \psi_2, \cdots, \psi_n$ von Linearformen derart, dass $|\psi_1|^2+\cdots+|\psi_n|^2 = |x_1|^2+\cdots+|x_n|^2$ wird. Sie liefern dann ein System mit unitärer Matrix $U=P+iQ$. In den reellen Veränderlichen u_ν, v_ν liefern sie $2n$ reelle Linearformen $\Psi_1, \cdots, \Psi_{2n}$ mit $\Psi_1=\Psi(u,v)$, die ein Orthogonalsystem mit der Koeffizientenmatrix $D=\begin{pmatrix} P & -Q \\ Q & P\end{pmatrix}$ bestimmen. Bildet man (vgl. 2) DRD^{-1} so enthält diese Matrix, weil $\Psi_1=\Psi(u,v)$ eine zu ω gehörende Invariante ist, in der ersten Zeile nur die Grössen $\omega, 0, 0, \cdots, 0$ und wegen der Symmetrie von R (und wegen $D^{-1}=D'$) gilt dasselbe auch für die erste Kolonne. Der Hilfssatz II lehrt uns aber, dass, wenn $UCU'=A^*+iB^*$ gesetzt wird, die erste Zeile von $DRD'=DRD^{-1}$ nur die Elemente der ersten Zeilen von A^* und B^* enthält. Hieraus folgt aber, dass die von uns gewonnene Matrix UCU' die Gestalt $UCU'=\begin{pmatrix} \omega & 0 & \cdots & 0 \\ 0 & c^*_{22} & \cdots & c^*_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & c^*_{n2} & \cdots & c^*_{nn}\end{pmatrix}$ hat. Das genügt aber, um den Satz I aussprechen zu können. Denn die Matrix der $c^*_{\kappa\lambda}$ $(\kappa, \lambda=2, 3\cdots n)$ ist wieder symmetrisch und kann in derselben Weise weiter behandelt werden, was zuletzt auf eine Schlussform $WCW'=\begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & \cdots & \cdots & \omega_n\end{pmatrix}$ mit nichtnegativem reellen ω_ν führt, bei unitärem W.

N. B. Hier ergibt sich, dass die ω eindeutig als die sämtlichen p positiven Wurzeln von $R=R(C)$ charakterisiert werden koennen, wozu noch $n-p$ Nullen hinzukommen. Man kann die $\omega_\nu{}^2$ aber auch als die Wurzeln der beiden „Normen" $\bar{C}C$ *und* $C\bar{C}$, die nichtnegative Hermitesche Formem liefern, kennzeichnen. Dies ergibt sich aus

$$C_1 = WCW', \; C_1\bar{C}_1 = WCW'\bar{W}\bar{C}\bar{W}' = WC\bar{C}W^{-1} = \begin{pmatrix} {\omega_1}^2 & 0 & \cdots & 0 \\ 0 & {\omega_2}^2 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & \cdots & {\omega_n}^2 \end{pmatrix};$$

analog für $\bar{C}C$ (vgl. die Bemerkung am Schluss der Notiz).

5. Folgerungen aus I. Wir wissen nunmehr, dass jede symmetrische Bilinearform $f = \sum\limits_{\kappa=1} c_{\kappa\lambda} x_\kappa y_\lambda$ sich in der folgenden Form screiben laesst:

$$f(x, y) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa y_\lambda = \omega_1 \phi_1(x)\, \phi_1(y) + \cdots + \omega_n \phi_n(x)\, \phi_n(y)$$

mit

$$\sum_{\nu=1}^{n} |\, \phi_\nu(x) \,|^2 = \sum_{\nu=1}^{n} |\, x_\nu \,|^2, \qquad \sum_{\nu=1}^{n} |\, \phi_\nu(y) \,|^2 = \sum_{\nu=1}^{n} |\, y_\nu \,|^2.$$

Die $\phi_1, \cdots, \phi_n$ kann man wegen der linearen Unabhängigkeit eines unitären Systems beliebig wählen, also auch so, dass $\phi_\nu = 1$, $\phi_\lambda = 0$ für $\lambda \neq \nu$ wird. Folglich sind die Werte $\omega_1, \omega_2, \cdots, \omega_n$ Werte der *quadratischen Form* $f(x, x)$ mit $|\, x_1 \,|^2 + \cdots + |\, x_\nu \,|^2 = 1$. Für $\sum\limits_{\nu=1}^{n} |\, x_\nu \,|^2 \leqq 1$, $\sum\limits_{\nu=1}^{n} |\, y_\nu \,|^2 \leqq 1$ erhalten wir ferner, wenn $\omega_1 \geqq \omega_2 \geqq \cdots \geqq \omega_n$ ist, nach der Schwarzschen Ungleichung

$$|\, f(x, y) \,| \leq \omega_1 \sum_{\nu=1}^{n} |\, \phi_\nu(x) \,| \; |\, \phi_\nu(y) \,|$$

$$\leq \omega_1 \sqrt{\sum_{\nu=1}^{n} |\, \phi_\nu(x) \,|^2 \sum_{\nu=1}^{n} |\, \phi_\nu(y) \,|^2} = \omega_1 \sqrt{\sum_{\nu=1}^{n} |\, x_\nu \,|^2 \sum_{\nu=1}^{n} |\, y_\nu \,|^2} \leq \omega_1.$$

Das gibt den für reelle $c_{\kappa\lambda}$ und x_ν, y_ν bekannten Satz:

III. *Jede symmetrische Bilinearform* $f(x, y)$ *hat die Eigenschaft, dass im Hilbertschen Raume* $\sum\limits_{\nu=1}^{n} |\, x_\nu \,|^2 \leq 1$, $\sum\limits_{\nu=1}^{n} |\, y_\nu \,|^2 \leq 1$ *das Maximum ihres absoluten Betrages auf dem Rande, und zwar schon für* $x_1 = y_1, \cdots, x_n = y_n$, *d. h. von der zugehörigen quadratischen Form* $f(x, x)$ *angenommen wird.*

Als unmittelbare Anwendungen ergeben sich Bemerkungen über quadratische Formen $\sum\limits_{\kappa,\lambda=1}^{\infty} c_{\kappa\lambda} x_\kappa x_\lambda$ $(c_{\kappa\lambda} = c_{\lambda\kappa})$ mit unendlich vielen Veraenderlichen, die im Hilbertschen Raume $\sum\limits_{\nu=1}^{\infty} |\, x_\nu \,|^2 \leq 1$ beschränkt sein sollen:

IV. *Eine quadratische Form* $\sum\limits_{\kappa,\lambda=1}^{\infty} c_{\kappa\lambda} x_\kappa x_\lambda$ *ist bei* $c_{\kappa\lambda} = c_{\lambda\kappa}$ *auch für komplexe* $c_{\kappa\lambda}$ *dann und nur dann beschränkt, wenn dies für die zugehörige*

Bilinearform der Fall ist. Quadratische Form und Bilinearform besitzen ferner die gleiche obere Schranke.

Es kommt weiter hinzu: Setzt man $c_{\kappa\lambda} = a_{\kappa\lambda} + i b_{\kappa\lambda}$ so ist die komplexe quadratische Form dann und nur dann beschränkt, wenn dies für die reelle quadratische Form $\sum\limits_{\kappa,\lambda=1}^{\infty} [a_{\kappa\lambda}(u_\kappa u_\lambda - v_\kappa v_\lambda) + b_{\kappa\lambda}(u_\kappa v_\lambda + v_\kappa u_\lambda)]$ in den Veränderlichen $u_1, v_1, u_2, v_2 \cdots$ gilt. Beim Uebergang von dem komplexen Fall zum reellen aendert sich die obere Schranke nicht.

Es ist also für $c_{\kappa\lambda} = c_{\lambda\kappa}$ nicht erforderlich, die "Normen" $C\bar{C}'$ bezw. $\bar{C}'C$ einzuführen.

Zusatz. Der Zusammenhang zwischen $R = \begin{pmatrix} A & B \\ B & -A \end{pmatrix}$ und den Normen

$$C\bar{C} = (A + iB)(A - iB)$$
$$= A^2 + B^2 + i(BA - AB) \text{ bezw. } \bar{C}C = A^2 + B^2 - i(BA - AB)$$

wird klar, wenn man $R^2 = \begin{pmatrix} A^2 + B^2, & AB - BA \\ BA - AB, & A^2 + B^2 \end{pmatrix}$ bildet. Diese Matrix hat die Gestalt $\begin{pmatrix} F & G \\ -G & F \end{pmatrix}$. Es wird aber

$$\begin{pmatrix} E & -iE \\ -iE & E \end{pmatrix} \begin{pmatrix} F & G \\ -G & F \end{pmatrix} = \begin{pmatrix} F + iG & 0 \\ 0 & F - iG \end{pmatrix} \begin{pmatrix} E & -iE \\ -iE & E \end{pmatrix}$$

wobei

$$\begin{pmatrix} E & -iE \\ -iE & E \end{pmatrix} \begin{pmatrix} E & iE \\ iE & E \end{pmatrix} = \begin{pmatrix} 2E & 0 \\ 0 & 2E \end{pmatrix}$$

wird. Dies zeigt, dass R^2 der Matrix $\begin{pmatrix} \bar{C}C & 0 \\ 0 & C\bar{C} \end{pmatrix}$ ähnlich ist. Die charakteristischen Wurzeln von $C\bar{C}$ (bezw. $\bar{C}C$) sind also in der Tat die Quadrate der Wurzeln von R, von zwei gleichen Quadraten nur eines gerechnet.

MEAN-VALUES OF ARITHMETICAL REPRESENTATIONS.*

By AUREL WINTNER.

By a representation, $f = f(n)$, will be meant any (number-valued) multiplicative representation of the ordinary multiplication on the semi-group of all positive integers; that is, any sequence of numbers $f(1), f(2), \cdots$ satisfying the relation $f(nm) = f(n)f(m)$ for every pair of positive integers n, m, distinct or not. Thus, if the trivial case $f(1) = 0$ (which leads to the identical vanishing of f) is excluded, then $f(1)$ must be 1, and so $f(n)$ becomes uniquely determined for every n by an arbitrary assignment of the values $f(p)$, where p runs through all primes.

It was recently shown ([2], p. 46) that every representation $f(n)$ which attains only the values ± 1 must have a mean

$$M(f) = \lim_{n \to \infty} \frac{f(1) + \cdots + f(n)}{n} . \tag{1}$$

The proof depended on *the prime number theorem* (which, in turn, follows by choosing $f(n) = \lambda(n)$, where λ denotes Liouville's function). In the present note, a parallel, but *elementary*, theorem will be verified by showing that every representation $f(n)$ which is bounded and non-negative has a mean (1). Here, and in the sequel, "non-negative" means "real and non-negative".

A class of functions more general than that of the representations results if the condition $f(nm) = f(n)f(m)$ is required only for integers n, m which are relatively prime. For these functions $f(n)$, the so-called multiplicative functions, not only the values $f(p)$, attained at primes, but all the values $f(p), f(p^2), \cdots$, attained at prime powers, can be chosen arbitrarily. The representations result if $f(p^k)$ is chosen to be the k-th power of the value $f(p)$, where p and k run through all primes and all positive integers respectively.

The theorem, quoted above concerning the case $f(n) = \pm 1$ of representations, was derived in [2], p. 46, in its more general form:

(i) *Every multiplicative function $f(n)$ which attains only the values ± 1 has a mean $M(f)$.*

---

* Received March 22, 1945.

Correspondingly, the elementary fact, stated above concerning representations, will be proved for the general case also:

(ii) *Every multiplicative function* $f(n)$ *which is non-negative and bounded has a mean* $M(f)$.

If $f(n) = \phi(n)/n$ or $f(n) = |\mu(n)|$, where ϕ and μ denote Euler's and Möbius' functions respectively, it is clear that the assumptions of (ii) are satisfied. Thus (ii) supplies the existence of (1) in both cases. In the first case, the existence of (1) goes back to Dirichlet (in fact, it is contained in his result concerning the sum-function of ϕ). In the second case, the existence of (1) means that the sequence of the square-free numbers has a density.

In both of these classical cases, the mere existence of (1) can of course be improved by *explicit* estimates of the error belonging to a fixed n. But (ii) is a *general* fact and cannot, therefore, supply improved remainder terms. For instance, (ii) contains the fact (cf., e. g., [1], p. 68) that every multiplicative set of positive integers has a density. But this particular case of (ii) is needed for the extension of the prime number theorem to the case of arbitrary sets of primes ([1], p. 71). In addition, this particular case of (ii) contains the truth of a conjecture of Ramanujan ([2], p. 31).

All of this seems to warrant an isolation of the general fact expressed by (ii).

In the proof, use will be made of the following elementary criterion: If $g(n)$ is a multiplicative function satisfying

$$(2)\qquad \sum_p \sum_{k=1}^{\infty} |g(p^k) - g(p^{k-1})|/p^k < \infty$$

(where $g(p^{k-1})$ is $g(1) = 1$ when $k = 1$), then the mean $M(g)$ exists; cf. [1], p. 41. It is also known that, if (2) is satisfied, the mean $M(g)$ of the multiplicative function g can be represented by the Eulerian product

$$(3)\qquad M(g) = \prod_p \left(1 + \sum_{k=1}^{\infty} \{g(p^k) - g(p^{k-1})\}/p^k\right);$$

cf. [1], p. 14.

What will actually be shown is that, if a multiplicative function $f(n)$ satisfies the assumptions of (ii) but is such that (2) fails to be satisfied for $g = f$, then $M(f)$ must exist and be 0. But it is known (cf. [1], p. 41) that, if a multiplicative function g satisfies (2), then g must be almost-periodic (B). On the other hand, if $M(f)$ exists and is 0, then, since $f \geqq 0$ by assumption, it is clear that f is almost-periodic (B), with $0 + 0 \cdots$ as Fourier series (B). Consequently, (ii) can be refined as follows:

(ii bis) *Every multiplicative function which is non-negative and bounded is almost-periodic* (B).

The possibility of this refinement of (ii) is due entirely to the *arithmetical* assumption of multiplicativity. In other words, if a function $f(n)$ is non-negative and bounded, then the mere existence of $M(f)$ does not imply the almost-periodicity (B) of $f(n)$. For a simple example proving more than the falsity of this implication, cf. [1], p. 28.

Let $f(n)$ be a function satisfying the assumptions of (ii). Thus

$$0 \leqq f(n) < A, \tag{4}$$

where A is a constant. In particular $|f(p^k)| < A$. Since $f(p^0) = f(1)$ is 1, it follows that (2) is satisfied for $g = f$ unless

$$\sum_p |f(p) - 1|/p = \infty. \tag{5}$$

Consequently, both (ii) and (ii bis) will be proved if it is shown that, if f satisfies the assumptions of (ii) and the additional assumption (5), then $M(f)$ exists and is 0.

To this end, let the sequence of all primes p be split into two complementary subsequences (either of which may be finite or vacuous), as follows: A prime p is in the first or in the second subsequence according as the value $f(p)$ is or is not less than 1. Thus, if a prime p is called a prime q or a prime r according as it is in the first or in the second subsequence,

$$(6_1) \qquad 0 \leqq f(q) < 1; \qquad\qquad (6_2) \qquad 1 \leqq f(r).$$

It is easy to see that

$$\sum_q |f(q) - 1|/q = \infty. \tag{7}$$

In fact, since q and r are complementary, it is clear from the assumption (5) that more than (7) is contained in the assertion

$$\sum_r |f(r) - 1| < \infty. \tag{8}$$

But (8) can be verified as follows:

Let $r_1, r_2, \cdots, r_j, \cdots$ denote the sequence of all primes r. It may be assumed that this sequence is infinite, since (8) is clear in the contrary case. For a fixed j, choose $n = r_1 r_2 \cdots r_j$ in (4). Then, since f is a multiplicative function, it follows that the partial products of the infinite product $\Pi f(r)$ form a bounded sequence. On the other hand, (6_2) shows that this sequence is non-decreasing. Accordingly, the infinite product $\Pi f(r)$ is convergent. In view of (6_2), this is equivalent to (8).

With reference to the given function $f(n)$ and to a fixed positive number t, let $f_t(n)$ denote the multiplicative function defined by the data $f_t(p^k)$ which are assigned as follows: If either p is an r or k is greater than 1, then $f_t(p^k)=f(p^k)$; while in the remaining case (that is, if p is a prime q and k is 1) the value of $f_t(p^k)=f_t(q)$ is $f(q)$ or 1 according as q is or is not less than t. Thus it is clear from (4) that

$$0 \leqq f_t(p^k) < A, \tag{9}$$

since A is greater than 1. Furthermore, it is seen from (6_1) and from the first of the inequalities (4) that the inequalities

$$0 \leqq f(n) \leqq f_t(n) \tag{10}$$

hold for every prime power, $n=p^k$. However, since both functions f, f_t are multiplicative, (10) must hold without the restriction $n=p^k$ also.

Condition (2) is satisfied by the function $g=f_t$ (for every fixed t). In fact, it is clear from (9) that the convergence of the double series (2) in the case $g=f_t$ is equivalent to the convergence of the simple series belonging to $k=1$. Since $f_t(1)=1$, it follows that (2) is equivalent to the convergence of

$$\sum_p |f_t(p)-1|/p=\sum_r |f_t(r)-1|/r+\sum_q |f_t(q)-1|/q.$$

But the convergence of the first series on the right is implied by (8), while the second series on the right has just a finite number of non-vanishing terms, since $f_t(q)$ was chosen to be 1 for every q exceeding t.

Since (2) is satisfied by $g=f_t$, it follows that $M(f_t)$ exists and is represented by the case $g=f_t$ of the absolutely convergent product (3). In view of the definition of $f_t(p^k)$, those factors of this infinite product which belong to primes r are independent of t. Hence, the product contributed by all these factors together can be denoted by C (without a subscript t). Then $M(f_t)=CQ_t$, where Q_t is an abbreviation for the product of all factors which belong to primes q.

Let Q_t be split into G_t times H_t, hence $M(f_t)$ into

$$M(f_t)=CG_tH_t, \tag{11}$$

where G_t comprises those factors of the product Q_t which belong to primes q having a value less than t, and H_t represents the contribution of all the remaining primes q. It is seen from the definition of $f_t(p^k)$ that, whether q

belongs to G_t or to H_t, the difference $f_t(q^k) - f_t(q^{k-1})$ is identical with $f(q^k) - f(q^{k-1})$ except when $k = 1$; and that, in the exceptional case $k = 1$, the difference $f_t(q^1) - f_t(q^0) = f_t(q) - 1$ is $f(q) - 1$ or $1 - 1 = 0$ according as q belongs to G_t or to H_t. If these data are substituted into the definitions of the contributions G_t, H_t to the case $g = f_t$ of the product (3), it follows that

$$G_t = \prod_{q<t} (1 + \{f(q) - 1\}/q + \sum_{k=2}^{\infty} \{f(q^k) - f(q^{k-1})\}/q^k) \tag{12}$$

and

$$H_t = \prod_{q \geqq t} (1 + \sum_{k=2}^{\infty} \{f(q^k) - f(q^{k-1})\}/q^k). \tag{13}$$

According to (12) and (4),

$$G_t = \prod_{q<t} (1 - \{1 - f(q)\}/q + O(q^{-2})).$$

It follows therefore from (6_1) that G_t is majorized by a constant multiple of

$$\prod_{q<t} (1 - \{1 - f(q)\}/q).$$

In view of (7) and (6_1), this means that $G_t \rightarrow 0$ as $t \rightarrow \infty$. Hence, in order to conclude from (11) that

$$M(f_t) \rightarrow 0 \quad \text{as} \quad t \rightarrow \infty, \tag{14}$$

it is sufficient to observe that, according to (13) and (4), the function H_t of t is majorized by the constant

$$\prod_{q} (1 + \sum_{k=2}^{\infty} 2A/q^k) < \infty.$$

Finally, it is clear from (14) and (10) that the limit (1) exists and is 0. But this is precisely what had to be proved.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Wintner, *Eratosthenian Averages*, Baltimore, 1943.
[2] ———, *The Theory of Measure in Arithmetical Semi-Groups*, Baltimore, 1944.

DIRICHLET SERIES MEROMORPHIC IN A HALF-PLANE.*

By H. D. Brunk.

1. Introduction and basic theorems.

The usefulness of Hadamard's famous "multiplication" theorem for Taylor series is well-recognized. Its direct analogue for Dirichlet series is very easily shown not to hold. However, Mandelbrojt has given powerful theorems ([4] and [5]; see also [6]) which generalize Hadamard's theorem for Dirichlet series. Generalizations of one of Mandelbrojt's theorems have been discussed in a recent paper of the author [2], and applications of these theorems and ideas form the subject matter of the present paper.

A function $f(s)$ is said to be represented by a Dirichlet series if

$$f(s) = \sum a_n e^{-\lambda_n s}, \tag{1}$$

where s is a complex variable, $s = \sigma + it$, $\{a_n\}$ is a sequence of complex constants, and $\{\lambda_n\}$ is a sequence of non-negative real numbers, strictly increasing to infinity. If the sequence $\{\lambda_n\}$ is the sequence of non-negative integers, the series is called a Taylor-Dirichlet series. Equation (1) is to be understood as implying that $f(s)$ is given by the series where the series converges, and is defined by analytic continuation wherever possible. Only uniform functions will be considered; this restriction is to be understood, whether or not it is specifically stated. However, the theorems of the paper apply to non-uniform functions as well, if cuts which render them uniform are allowed to play the rôles otherwise assigned to singular points of uniform functions. Also, only those functions will be considered here which are represented by Dirichlet series having finite abscissae of absolute convergence. V. Bernstein's notation will be used: $\mathcal{A}_f$ denotes the abscissa of absolute convergence of the function $f(s)$, and $\mathcal{H}_f$ denotes its abscissa of holomorphism.

The definitions of the next few paragraphs will be required in order to state the fundamental theorem, Theorem 1. The symbol $+$, as an operator on sets of points, will be used as it is commonly used in point set theory; so that if A and B are sets, $A + B$ is their sum, or union. Also $A \cdot B$ denotes their intersection, the set of all points common to both. The word curve, when

---

* Received August 27, 1944; Presented to the American Mathematical Society April 28, 1944.

used, refers to a simple Jordan curve. The term domain will refer to an open, connected set of points; the term region will be used in referring to a domain plus all, some, or none of its boundary points. A point will be considered singular for a function if it is not possible to continue the function analytically throughout a complete neighborhood of the point.

DEFINITION. A domain O is *bounded connectedly with respect to a domain* D if there exists a positive number K such that to every point of O corresponds a simple Jordan curve in $O + D$ of length less than K which joins this point to points in D. It is of course implied that $O + D$ is itself a domain.

Let $f(s)$ be given by (1), and let S_f denote the set of singular points of $f(s)$ in the complex s-plane, where $s = \sigma + it$. Let c be a real number greater than $\mathfrak{a}_f$, and let A be a closed set in $\sigma < c$ containing S_f.

DEFINITION. *The function* $f(s)$ *is "M" in* $\sigma > \sigma_1$ *with respect to* A *if* $f(s)$ *is bounded in every domain which is bounded connectedly with respect to* $\sigma > c$ *and whose boundary is at positive distance from the set* $A + P_1$, *where* P_1 *is the half-plane* $\sigma \leq \sigma_1$. Since $f(s)$ is bounded in the half-plane $\sigma \geq c$ if $c > \mathfrak{a}_f$, the definition is clearly independent of the choice of the real number c greater than $\mathfrak{a}_f$. *If* $f(s)$ *satisfies the above condition for every real number* σ_1, *then* $f(s)$ *is "M" in the entire plane with respect to* A. *If* $f(s)$ *satisfies the condition with* A *replaced by* S_f, *then* $f(s)$ *is "M" in* $\sigma > \sigma_1$.

The function $f(s)$ is "*bounded except for singularities*" in $\sigma > \sigma_1$ if $f(s)$ is bounded in that part of the half-plane $\sigma > \sigma_1$ remaining after circles of arbitrary, fixed radius about points of S_f have been extracted. It has been observed [2] that *if a function* $f(s)$ *is bounded except for singularities in a certain half-plane, it is also "M" in that half-plane.* The converse is not true. It has also been shown [2] that *if* $f(s)$ *fails to take two distinct values* a *and* b *in a strip* $\sigma_1 < \sigma < c$ *for real numbers* σ_1 *and* c, *where* $c > \mathfrak{a}_f$, *then* $f(s)$ *is "M" in* $\sigma > \sigma_1$. In fact, *this is true if* $f(s)$ *takes at most a finite number of times in* $\sigma_1 < \sigma < c$ *each of two values* a *and* b. For it was seen (Lemma 1 of [2]) that $f(s)$ is bounded in every domain bounded connectedly with respect to $\sigma > c$ which is at positive distance from $A + P_1$, where $A = S_f + E_f(a) + E_f(b)$, and $P_1 = (\sigma \leq \sigma_1)$; here $E_f(a)$ denotes the set of points in $\sigma_1 < \sigma < c$ at which $f(s)$ takes the value a, and $E_f(b)$ the set in $\sigma_1 < \sigma < c$ on which $f(s)$ takes the value b. Suppose that at most a finite number of points is taken from each of the sets $E_f(a)$ and $E_f(b)$ to obtain the sets $E'_f(a)$ and $E'_f(b)$. Let O' be any domain bounded connectedly with respect to $\sigma > c$ which is at positive distance from $S_f + E'_f(a) + E'_f(b) + P_1$. If O' contains any points of $E_f(a) + E_f(b)$, extract from O' closed, non-overlapping circles interior to O' about each of these points; let O be the domain remaining. Then O is bounded connectedly with respect to $\sigma > c$ and is at positive distance from each point of $S_f + E_f(a) + E_f(b) + P_1$, so that $f(s)$ is bounded in O. Also, $f(s)$ is bounded in each of the extracted circles, which are closed, bounded sets in each of which $f(s)$ is holomorphic. Therefore $f(s)$ is bounded in O'. By definition, then, $f(s)$ is "M" in $\sigma > \sigma_1$ with respect to $S_f + E'_f(a) + E'_f(b)$. If $E_f(a) + E_f(b)$ contains at most a finite number of points, the above discussion applies with $E'_f(a)$ and $E'_f(b)$ both empty, so that $f(s)$ is "M" in $\sigma > \sigma_1$.

Lemma 1 of [2], referred to above, may be regarded as stating that *any function* $f(s)$ *given by a Dirichlet series having a finite abscissa of absolute convergence is* "M" *in* $\sigma > \sigma_1$ *with respect to* $S_f + E_f(a) + E_f(b)$, *where* a *and* b *are any two distinct complex numbers.*

S. Mandelbrojt ([4]; page 4) has given examples of functions bounded except for singularities, which are therefore also "M." Examples of "M" functions which are not bounded except for singularities are given in [2].

Consider two functions

$$f(s) = \sum a_n e^{-\lambda_n s}, \qquad \phi(s) = \sum b_n e^{-\lambda_n s}, \tag{2}$$

and the composite function

$$H(f, \phi) = H(s) = \sum a_n b_n e^{-\lambda_n s}. \tag{3}$$

Let S_ϕ denote the set of singular points of $\phi(s)$. Let c and c' be real numbers greater than $\mathfrak{A}_f$ and $\mathfrak{A}_\phi$ respectively. Let A be a closed set in $\sigma < c$ containing S_f, and B a closed set in $\sigma < c'$ containing S_ϕ. Define

$$\begin{aligned}
\sigma(A) &= \max\,[\sigma_1, \underset{\sigma+i\tau\in A,\ -\infty<\tau<\infty}{\text{l. u. b.}} \sigma], \\
\sigma^\tau(A) &= \max\,[\sigma_1, \underset{\sigma+i\tau\in A}{\text{l. u. b.}}\ \sigma], \\
\sigma^\tau(B) &= \max\,[\sigma_1, \underset{\sigma+i\tau\in B}{\text{l. u. b.}}\ \sigma], \\
\sigma^\infty(A) &= \max\,[\limsup_{\tau\to\infty} \sigma^\tau(A), \limsup_{\tau\to-\infty} \sigma^\tau(A)], \\
\sigma^\infty(B) &= \max\,[\limsup_{\tau\to\infty} \sigma^\tau(B), \limsup_{\tau\to-\infty} \sigma^\tau(B)], \\
\sigma^\infty &= \max\,[\sigma_1 + \sigma^\infty(B), \sigma_2 + \sigma^\infty(A)].
\end{aligned} \tag{4}$$

Let $[A, B]$ denote the closure of the set of all points $\alpha + \beta$, where $\alpha \in A$, $\beta \in B$. It is, by definition, a closed set. Let A_τ denote the intersection of A

with the quarter-plane $t \geqq \tau$, $\sigma > \sigma_1$ of the complex s-plane, where $s = \sigma + it$: $A_\tau = A \cdot (t \geqq \tau, \sigma > \sigma_1)$. Define

$$(5) \qquad [A, B]^\infty = \lim_{\tau \to \infty} [A_\tau, B],$$

the intersection of all sets $[A_\tau, B]$ as τ takes on all real values, which is itself a closed set (see [2]).

The theorem (cf. Theorem 2 of [2]) on which most of the theorems of the present paper are based may now be stated:

THEOREM 1. *Let A be any closed set in $\sigma < c$ which contains S_f, and let B be any closed set in $\sigma < c'$ which contains S_ϕ. If $f(s)$ is "M" in $\sigma > \sigma_1$ with respect to A, and if $\phi(s)$ is "M" in $\sigma > \sigma_2$ with respect to B, then $H(s)$ is holomorphic in that part of the complement in $\sigma > \sigma^\infty$ of $[A, B]^\infty \cdot (\sigma > \sigma^\infty)$ which is connected with $\sigma > c + c'$. Moreover, $H(s)$ is "M" in $\sigma > \sigma^\infty$ with respect to $[A, B]^\infty$.*

Proof. The theorem as stated above is precisely Theorem 2 of [2], with the exception of the last sentence, which is not included there. It remains, then, only to prove that under the given conditions $H(s)$ is "M" in $\sigma > \sigma^\infty$ with respect to $[A, B]^\infty$.

The sets $A(\delta)$ and $B(\delta)$ will be defined as in [2]: $A(\delta)$ is the complement in $\sigma \geqq \sigma_1$ of the set $CA(\delta)$ defined as follows. Let $P(\delta)$ be a positive function of the positive variable δ, monotonely increasing to infinity as δ approaches zero. The set $CA(\delta)$ is the set of all curves of length less than $P(\delta)$, having one end-point in $\sigma > c$, which are central lines of channels in $\sigma > \sigma_1$ of width greater than 2δ which contain no points of A. $B(\delta)$ is defined in a similar manner, replacing σ_1 by σ_2, c by c', and A by B.

Let T denote the closed set $[A, B]^\infty$, which lies in the half-plane $\sigma \leqq c + c'$. Define $T(\delta)$ as $A(\delta)$ was defined, replacing σ_1 by σ^∞, c by $c + c'$, and A by T. It is clear that every domain which is bounded connectedly with respect to $\sigma > c + c'$ and which is at positive distance from the set $T + P$, where P is the half-plane $\sigma \leqq \sigma^\infty$, is interior to a domain $CT(\delta)$ for some positive δ. Also, every domain $CT(\delta)$ is itself bounded connectedly with respect to $\sigma > c + c'$, and is at positive distance from the set $T + P$. In order to show that $H(s)$ is "M" in $\sigma > \sigma^\infty$ with respect to T, it is therefore sufficient to show that $H(s)$ is bounded in each domain $CT(\delta)$.

Let δ' and δ'' be positive numbers such that $\delta' + \delta'' < \delta$, and write $\epsilon = [\delta - (\delta' + \delta'')]/5$. Let $M_f(\delta')$ be the least upper bound of $|f(s)|$ in $CA(\delta')$, and $M_\phi(\delta'')$ the least upper bound of $|\phi(s)|$ in $CB(\delta'')$. It can easily be seen, by considering Mandelbrojt's integral expression for $H(s)$ (see [2], Theorem 1) that $H(s)$ is bounded in $CT(\delta)$. If $M_H(\delta)$ denotes the least upper bound of $|H(s)|$ in $CT(\delta)$, it can be shown that actually there exists a constant N independent of δ, δ', and δ'', such that

$$(6) \qquad M_H(\delta) = M_H(5\epsilon + \delta' + \delta'') < (N/\epsilon) M_f(\delta') M_\phi(\delta'').$$

In order to show this, let z' be any point of $CT(\delta)$. There is then a curve L of length less than $P(\delta)$ joining z' to points in $\sigma > c + c'$, which is the central line of a channel in $\sigma > \sigma^\infty$ of width greater than 2δ which contains no point of $[A, B]^\infty$. It follows from Lemma 2 of [2] that if δ is sufficiently small, and if $\tau > \tau_0$ for τ_0 sufficiently large, then L is also the central line of a channel C' of width greater than $2[\delta - (\delta' + \delta'')]$, or 10ϵ, in $\sigma > \sigma^\infty$, which contains no point of $[A_\tau(\delta'), B(\delta'')]$, where $A_\tau(\delta') = A(\delta') \cdot (t \geq \tau)$. There is therefore a domain Δ in $\sigma > \sigma^\infty + 4\epsilon$ containing L and also points with arbitrarily large real part, whose boundary is at distance greater than 4ϵ from every point of $[A_\tau(\delta'), B(\delta'')]$, for $\tau > \tau_0$, and for δ sufficiently small. Fix δ', δ'', δ, and τ, so that this is true.

Pave the strip $\sigma_1 \leq \sigma \leq c$ of the s-plane ($s = \sigma + it$) with squares of side ϵ. Extract from this strip those closed squares, as well as those bordering on them, which contain points of $A_{\tau_0}(\delta')$. Extract also those bordering on the vertical line $\sigma = \sigma_1$. Denote by D_ϵ that part of the region in the half-plane $\sigma > \sigma_1$ remaining which is connected with the half-plane $\sigma > c$. Define the sequence $\{\alpha_m\}$ and the symbol $C_\epsilon(\alpha_m)$ as in Theorem 1 of [2]. It follows from the discussion in the proof of that theorem that

$$H(z') = \lim_{m \to \infty} (1/i\alpha_m) \int_{C_\epsilon(\alpha_m)} f(s) \phi(z' - s)\, ds.$$

Also, $C_\epsilon(\alpha_m)$ is interior to $CA(\delta')$, so that on $C_\epsilon(\alpha_m)$, $|f(s)| \leq M_f(\delta')$; and for s on $C_\epsilon(\alpha_m)$, $z' - s$ is in $CB(\delta'')$, so that $|\phi(z' - s)| \leq M_\phi(\delta'')$. It is pointed out in [2] that there is a constant N independent of ϵ such that the length of $C_\epsilon(\alpha_m)$ is less than $N\alpha_m/\epsilon$. It follows that

$$|H(z')| < (1/\alpha_m) M_f(\delta') M_\phi(\delta'') (N\alpha_m/\epsilon),$$

$$|H(z')| < (N/\epsilon) M_f(\delta') M_\phi(\delta'').$$

But z' may be any point of $CT(\delta)$, so that

$$M_H(\delta) = M_H(5\epsilon + \delta' + \delta'') < (N/\epsilon) M_f(\delta') M_\phi(\delta'');$$

which completes the proof of the theorem.

Mandelbrojt (see [6]) has introduced the concept "ultimate positive abscissa of holomorphism of $f(s)$ with respect to σ_1." This real number is the larger of the two numbers σ_1 and the upper limit as τ becomes infinite of the largest real number σ such that $\sigma + i\tau \in S_f$, where S_f denotes the set of singular points of $f(s)$. A closely allied terminology will be used here. The *ultimate abscissa of holomorphism of* $f(s)$, $\mathcal{H}^\infty(f)$, is defined as follows:

$$\mathcal{H}^\infty(f) = \max \, [\limsup_{\tau \to -\infty} (\max_{\sigma + i\tau \in S_f} \sigma), \ \limsup_{\tau \to \infty} (\max_{\sigma + i\tau \in S_f} \sigma)]. \tag{7}$$

It will be noted that with the previously used notation,

$$\sigma^\infty(S_f) = \max \, [\sigma_1, \mathcal{H}^\infty(f)].$$

THEOREM 2. *If* $f(s) = \sum a_n e^{-\lambda_n s}$ *is "M" in* $\sigma > \sigma_0$ *for some real number* σ_0 *less than* $\mathcal{H}_f - \limsup (\log n)/\lambda_n$, *then* $\mathcal{H}^\infty(f) \geq \mathcal{H}_f - \limsup (\log n)/\lambda_n$.

In particular, since $\mathcal{H}^\infty(f) \leq \mathcal{H}_f$, if $\limsup (\log n)/\lambda_n = 0$, and if $f(s)$ is "M" in $\sigma > \sigma_0$ for some real number σ_0 less than $\mathcal{H}_f$, then $\mathcal{H}^\infty(f) = \mathcal{H}_f$; that is, $f(s)$ has singular points, the absolute values of whose imaginary parts are arbitrarily large, in every open half-plane containing the half-plane $\sigma \geq \mathcal{H}_f$ in its interior.

Proof. Suppose that the theorem is not true, i.e., that $\mathcal{H}^\infty(f) < \mathcal{H}_f - \limsup (\log n)/\lambda_n$. Choose σ_1 so that $\sigma_0 < \sigma_1 < \mathcal{H}_f - \limsup (\log n)/\lambda_n$, and $\sigma_1 > \mathcal{H}^\infty(f)$. Then $f(s)$ is "M" in $\sigma > \sigma_1$, and $\sigma^\infty(S_f) = \max(\sigma_1, \mathcal{H}^\infty(f)) = \sigma_1$. Put $A = S_f$ in Theorem 1; then A_τ, or $S_f \cdot (t \geq \tau, \sigma > \sigma_1)$, is empty for τ sufficiently large, and $[A, B]^\infty$ is empty, whatever the set B may be. In Theorem 1, put $\phi(s) = \sum e^{-\lambda_n s}$. Choose the set B as indicated in Section 5 of [2], so that $\phi(s)$ is "M" in the entire plane with respect to B. Since

$$\mathcal{A}_\phi = \limsup \, (\log n)/\lambda_n \tag{8}$$

(see [1] or [7]), such a set B can be constructed as follows: let $\{t_k\}$ be any sequence of non-negative numbers, strictly increasing to infinity, such that $\lim_{k\to\infty} (t_{k+1} - t_k) = 0$. Define $B = S_\phi + B_1$, where S_ϕ is the set of all singular points of $\phi(s)$, and B_1 is the set of points $s_k = \mathcal{A}_\phi + it_k$ and $s'_k = \mathcal{A}_\phi - it_k$ $(k = 1, 2, \cdots)$. Then $\phi(s)$ is "M" in $\sigma > \sigma_1$ with respect to B for $-\sigma_2$ arbitrarily large. The number σ_2 can be chosen so that $\sigma^\infty = \max \, [\sigma_1 + \sigma^\infty(B), \sigma_2 + \sigma^\infty(A)] = \sigma_1 + \sigma^\infty(B) = \sigma_1 + \mathcal{A}_\phi = \sigma_1 + \limsup (\log n)/\lambda_n$. Since $\sigma_1 < \mathcal{H}_f - \limsup (\log n)/\lambda_n$, $\sigma^\infty < \mathcal{H}_f$. But $[S_f, B]^\infty$ is empty, so that it follows from Theorem 1 that $H(s) = \sum a_n e^{-\lambda_n s} = f(s)$ is holomorphic in $\sigma > \sigma^\infty$. This evidently contradicts the above statement that $\mathcal{H}_f > \sigma^\infty$, and completes the proof of the theorem.

THEOREM 3. *If* $f(s) = \Sigma a_n e^{-\lambda_n s}$ *is "M" in* $\sigma > \sigma_1$ *for some real number* $\sigma_1 < \mathcal{A}_f - \limsup (\log n)/\lambda_n$, *then* $\mathcal{H}_f \geqq \mathcal{A}_f - \limsup (\log n)/\lambda_n$.

Proof. Suppose that the theorem is not true, i. e., that $\mathcal{H}_f < \mathcal{A}_f - \limsup (\log n)/\lambda_n$. Let d be any real number less than $\mathcal{A}_f \limsup (\log n)\lambda_n$ and greater than both σ_1 and $\mathcal{H}_f$. Since $\sigma_1 < d$, and since $f(s)$ has no singular points in $\sigma \geqq d$, it follows from the definition of an "M" function that $f(s)$ is bounded by some positive number N in $\sigma \geqq d$. Use of this fact with Cauchy's theorem in a classical manner yields the result that

$$\lim_{T\to\infty} (1/iT) \int_{d+0i}^{d+iT} f(s) e^{\lambda_n s}\, ds = \lim_{T\to\infty} (1/iT) \int_{c+0i}^{c+iT} f(s) e^{\lambda_n s}\, ds,$$

where c is a real number greater than $\mathcal{A}_f$, and the path of integration in each case is the straight line joining the end-points. But, as is well-known (e. g. [7], page 15),

$$a_n = \lim_{T\to\infty} (1/iT) \int_{c+0i}^{c+iT} f(s) e^{\lambda_n s}\, ds.$$

Also,

$$| a_n | = | \lim_{T\to\infty} (1/iT) \int_{d+0i}^{d+iT} f(s) e^{\lambda_n s}\, ds | \leqq N e^{\lambda_n d}.$$

Hence

$$\Sigma \,| a_n e^{-\lambda_n z} | \leqq \Sigma \,| a_n | \exp[-\lambda_n \mathcal{R}(z)] \leqq N \Sigma \exp\{-\lambda_n [\mathcal{R}(z) - d]\}.$$

But by (8), $\Sigma e^{-\lambda_n s}$ converges if $\mathcal{R}(s) > \limsup (\log n)/\lambda_n$. Thus if $\mathcal{R}(z) > d + \limsup (\log n)/\lambda_n$, $\Sigma a_n e^{-\lambda_n z}$ converges absolutely. Therefore $\mathcal{A}_f \leqq d + \limsup (\log n)/\lambda_n$. But $d < \mathcal{A}_f - \limsup (\log n)\lambda_n$, a contradiction which completes the proof of the theorem.

COROLLARY 1. *If* $f(s) = \Sigma a_n e^{-\lambda_n s}$ *is "M" in an open half-plane containing in its interior its axis of absolute convergence, and if* $\lim (\log n)/\lambda_n = 0$, *then* $\mathcal{H}^{\infty}(f) = \mathcal{H}_f = \mathcal{A}_f$.

COROLLARY 2. *If* $f(s)$ *is any function given by a Dirichlet series having a finite abscissa of absolute convergence, and if* $\sigma_1 < \mathcal{A}_f - 2 \limsup (\log n)/\lambda_n$ (*or if* $\sigma_1 < \mathcal{H}_f - \limsup (\log n)/\lambda_n$) *then either*

(i) $f(s)$ *has singular points, the absolute values of whose imaginary parts are arbitrarily large, in* $\sigma > \sigma_1$, *or*

(ii) $f(s)$ *takes infinitely often every value except at most one in* $\sigma_1 < \sigma < c$, *where* c *is any real number greater than* $\mathcal{A}_f$.

For if (ii) is not satisfied, then $f(s)$ is "M" in $\sigma > \sigma_1$, as is noted near the beginning of this section. Then by Theorem 3, $\mathcal{H}_f \geq \mathcal{A}_f - \limsup (\log n)/\lambda_n$, so that $\sigma_1 < \mathcal{H}_f - \limsup (\log n)/\lambda_n$. By Theorem 2, $\mathcal{H}^\infty(f) \geq \mathcal{H}_f - \limsup (\log n)/\lambda_n > \sigma_1$, so that (i) is satisfied.

Two lemmas, which will be used in developing applications of the above theorems, follow.

LEMMA 1. *Let* $\{f_i(s)\}$ *be a finite set of functions,*

$$f_i(s) = \sum a_n^{(i)} \exp\{-\lambda_n^{(i)} s\}, \qquad (i = 1, 2, \cdots, N),$$

which satisfy the following conditions:

(i) $\max_{i=1,2,\ldots,N} \mathcal{A}_{f_i} = \mathcal{A} < \infty$;

(ii) $\lambda_m^{(i)} \neq \lambda_n^{(j)}$ *for all* m *and* n, *if* $i \neq j$;

(iii) $\lim (\log n)/\mu_n = 0$, *where the sequence* $\{\mu_n\}$ *is composed of the elements of all the sequences* $\{\lambda_n^{(i)}\}$, *arranged in increasing order;*

(iv) $F(s) = \sum_{i=1}^{N} f_i(s) = \sum b_n e^{-\mu_n s}$ *is "M" in* $\sigma > \sigma_0$ *for some real number* $\sigma_0 < \mathcal{A}$.

Then $\mathcal{H}_F = \mathcal{A}_F = \mathcal{A}$.

Proof. Under the hypotheses of the lemma (see [1] or [7]),

$$\mathcal{A}_F = \limsup \frac{\log |b_n|}{\mu_n} = \max_{i=1,2,\ldots,N} \limsup_{n\to\infty} \frac{\log |a_n^{(i)}|}{\lambda_n^{(i)}} = \mathcal{A}.$$

But by Corollary 1, $\mathcal{H}_F = \mathcal{A}_F$, which completes the proof.

LEMMA 2. *Suppose* $f(s) = \sum a_n e^{-\lambda_n s}$ *and* $\gamma(s) = \sum b_n e^{-\mu_n s}$ *are both "M" in* $\sigma > \sigma_1$ *for some real number* σ_1. *Let* $F(s)$ *denote their sum:* $F(s) = f(s) + \gamma(s)$. *Let* E *be the set of points (if any) in* $\sigma > \sigma_1$ *which are singular for both* $f(s)$ *and* $\gamma(s)$ *and at which* $F(s)$ *is holomorphic. If* E *is empty, or if the greatest lower bound of the distances between pairs of points of* E *is positive, then* $F(s)$ *is "M" in* $\sigma > \sigma_1$.

Proof. Let A denote the set S_f of singular points of $f(s)$ in $\sigma > \sigma_1$ and B the set S_γ of singular points of $\gamma(s)$ in $\sigma > \sigma_1$. Under the hypotheses, there exists a positive function $M_f(\delta)$ of the positive variable δ such that $|f(s)| < M_f(\delta)$ in $CA(\delta)$, which is defined near the beginning of this section. Similarly, there exists a function $M_\gamma(\delta)$ such that $|\gamma(s)| < M_\gamma(\delta)$ in $CB(\delta)$, defined as is $CA(\delta)$, replacing A by B, and c, a number greater than $\mathfrak{a}_F$, by c_1, a number greater than $\mathfrak{a}_\gamma$. Let T be the set S_F of singular points of $F(s)$ in $\sigma > \sigma_1$, and define $CT(\delta)$ as $CA(\delta)$ was defined, replacing A by T, and c by $\max(c, c_1)$, which is certainly greater than $\mathfrak{a}_F$.

Let d be a positive number such that the distance between any two points of E is greater than d; let δ be any positive number less than $\frac{1}{4}d$, and let x be any point of $CT(\delta)$. There is then a curve L of length less than $P(\delta)$ (hence less than $P(\frac{1}{4}\delta)$) which joins x to $\sigma > \max(c, c_1)$ and is the central line of a channel C in $\sigma > \sigma_1$ of width greater than 2δ which contains no point of S_F. It can therefore contain no points of S_f or of S_γ except possibly points of E. There are at most a finite number n, depending only on δ, of points y_i $(i = 1, 2, \cdots, n)$ of E in C; for any channel whose central line is of length less than $P(\delta)$ can be covered by a finite number of circles of radius $\frac{1}{2}d$; each such circle contains at most one point of E. If y_i is any point of E, let K_i be a circle of radius $\frac{1}{4}\delta$ about y_i. Since $\delta < \frac{1}{4}d$, no two such circles intersect. If each K_i fails to intersect L, then L is the central line of a channel in $\sigma > \sigma_1$ of width greater than $\frac{1}{4}\delta$ which contains no point of A or B, so that x belongs to both $CA(\frac{1}{4}\delta)$ and $CB(\frac{1}{4}\delta)$, and $|f(x)| < M_f(\frac{1}{4}\delta)$, $|\gamma(x)| < M_\gamma(\frac{1}{4}\delta)$, therefore $|F(x)| < M_f(\frac{1}{4}\delta) + M_\gamma(\frac{1}{4}\delta)$.

If a circle K_i intersects L, K_i is at distance greater than $\frac{1}{2}\delta$ from the boundary of C.

If no K_i contains x, L may be replaced by a curve L', which is obtained by replacing every arc (if any) of L which is in a circle K_i by an arc of K_i between the two points of intersection. The length of L' is certainly less than $P(\delta) + 2n\pi\delta/4$, which is less than $P(k\delta)$ for some positive number k less than $1/8$, depending only on δ, since $P(\delta)$ increases monotonely to infinity as δ approaches zero. Moreover, L' joins x to $\sigma > \max(c, c_1)$, and is the central line of a channel in $\sigma > \sigma_1$ of width greater than $k\delta$ $(k < 1/8)$ which contains no point of A or B. Therefore $x \in CA(k\delta)$ and $x \in CB(k\delta)$. Hence $|F(x)| < M_f(k\delta) + M_\gamma(k\delta)$.

Finally, if a circle K_j contains x, there is only one such circle. By the discussion of the paragraph above, each point x' on the boundary of K_j lies in $CA(k\delta)$ and $CB(k\delta)$. Therefore $|F(x')| < M_f(k\delta) + M_\gamma(k\delta)$ for each such point. But $F(s)$ is holomorphic in K_j, and therefore takes its maximum value on the boundary, so that $|F(x)| < M_f(k\delta) + M_\gamma(k\delta)$.

This last inequality has been shown to hold for each point x in $CT(\delta)$, for each positive δ (k depending on δ). Thus $F(s)$ is bounded in each domain $CT(\delta)$; as has been noted before, this is sufficient in order that $F(s)$ be "M" in $\sigma > \sigma_1$. This completes the proof of the lemma.

The generalization for the sum of a finite number of functions is immediate. If each of the functions is "*M*," if E is the set of points singular for some or all of the functions on which the sum is holomorphic, and if the greatest lower bound of the distances between pairs of points of E is positive, then the sum also is "*M*."

2. Applications. In this section, theorems based on the theorems and lemmas of the previous section will be proved. First, it will have been noticed that the composite function $H(f,\phi)$ of Theorem 1 is defined only if the functions $f(s)$ and $\phi(s)$ are of the form: $f(s)=\Sigma a_n e^{-\lambda_n s}$, $\phi(s)=\Sigma b_n e^{-\lambda_n s}$, the two series having the same sequence of exponents. If, however, $f(s)=\Sigma a_n e^{-\lambda_n s}$ and $\phi(s)=\Sigma b_n e^{-\mu_n s}$, there is a minimum sequence $\{\nu_n\}$ which contains both sequences, $\{\lambda_n\}$ and $\{\mu_n\}$. The functions $f(s)$ and $\phi(s)$ can each be represented by a Dirichlet series whose sequence of exponents is $\{\nu_n\}$:

$$f(s)=\sum A_m e^{-\nu_m s};\qquad A_m=a_n \quad\text{if}\quad \nu_m=\lambda_n;$$

$$A_m=0 \quad\text{if}\quad \nu_m\neq\lambda_n \text{ for all } n;$$

$$\phi(s)=\sum B_m e^{-\nu_m s};\qquad B_m=b_n \quad\text{if}\quad \nu_m=\mu_n;$$

$$B_m=0 \quad\text{if}\quad \nu_m\neq\mu_n \text{ for all } n.$$

If the sequences $\{\lambda_n\}$ and $\{\mu_n\}$ have an infinite sequence $\{r_n\}=\{\lambda_n\}\cdot\{\mu_n\}$ in common ($\{r_n\}$ is the sequence of all non-negative numbers, arranged in increasing order, each of which appears in both $\{\lambda_n\}$ and $\{\mu_n\}$), if $r_i=\lambda_{n_i}=\mu_{m_i}$ $(i=1,2,\cdots)$, and if

$$f_1(s)=\sum a_{n_i}e^{-\lambda_{n_i}s},\qquad \phi_1(s)=\sum b_{m_i}e^{-\mu_{m_i}s},$$

then

$$H(f,\phi)=H(f_1,\phi_1)=\sum A_m B_m e^{-\nu_m s}=\sum a_{n_i}b_{m_i}e^{-r_i s}.$$

If $\{\lambda_n\}$ and $\{\mu_n\}$ have no infinite sequence in common, the result obtained by the use of Theorem 1 is of course trivial, since then $H(f,\phi)$ is entire or constant. Use will be made in this section of Theorem 1, applied to functions whose sequences of exponents are not identical, but have an infinite subsequence in common. One such result follows, indicating that under certain conditions, two "*M*" functions represented by Dirichlet series can not each "cancel" the other's singularities so that their sum is holomorphic, unless they have essentially the same sequences of exponents.

THEOREM 4. *Suppose that* $F(s) = \Sigma A_n e^{-\lambda_n s}$ *and* $\Phi(s) = \Sigma B_n e^{-\mu_n s}$ *satisfy the following conditions:*

(i) $F(s)$ *is "M" in* $\sigma > \mathfrak{A}_F - \eta$, $\Phi(s)$ *is "M" in* $\sigma > \mathfrak{A}_\Phi - \eta$, *for some positive number* η;

(ii) $F(s) - \Phi(s)$ *is holomorphic and "M" in* $\sigma > \mathfrak{A}_F - \eta$;

(iii) $\lim (\log | A_n |)/\lambda_n = \mathfrak{A}_F$,

$\lim (\log | B_n |)/\mu_n = \mathfrak{A}_\Phi$;

(iv) $\lim (\log n)/\lambda_n = \lim (\log n)/\mu_n = 0$;

Then the sequences $\{\lambda_n\}$ *and* $\{\mu_n\}$ *are identical, with the exception of at most a finite number of terms of each sequence.*

It is evident from Corollary 1 of Theorems 2 and 3 that under hypotheses (i) and (iv), $\mathfrak{H}_\Phi = \mathfrak{A}_\Phi$ and $\mathfrak{H}_F = \mathfrak{A}_F$. It follows from hypothesis (ii) that $\mathfrak{A}_F = \mathfrak{A}_\Phi$.

Incidentally, it will be observed that Lemma 2 assures the fulfillment of condition (ii), if (i) is satisfied, if $F(s) - \Phi(s)$ is holomorphic in $\sigma > \mathfrak{A}_F - \eta$, and if further the greatest lower bound of the distances between pairs of points, singular for both $F(s)$ and $\Phi(s)$ and at which $F(s) - \Phi(s)$ is holomorphic, is positive.

Proof. It will first be shown that all of the terms of the sequence $\{\lambda_n\}$ except at most a finite number belong also to the sequence $\{\mu_n\}$. Write

$$f(s) = F(s + \mathfrak{A}_F) = \sum A_n \exp\,[-\lambda_n(s + \mathfrak{A}_f)] = \sum a_n e^{-\lambda_n s},$$

where

$$a_n = A_n \exp\,[-\lambda_n \mathfrak{A}_F];$$

and

$$\phi(s) = \Phi(s + \mathfrak{A}_F) = \sum B_n \exp\,[-\mu_n(s + \mathfrak{A}_F)] = \sum b_n e^{-\mu_n s},$$

where

$$b_n = B_n \exp\,[-\lambda_n \mathfrak{A}_F].$$

Then the hypotheses of Theorem 4 become:

(i) $f(s)$ is "M" in $\sigma > -\eta$; $\phi(s)$ is "M" in $\sigma > -\eta$; $\mathfrak{A}_F = \mathfrak{A}_\Phi = 0$;

(ii) $f(s) - \phi(s)$ is holomorphic and "M" in $\sigma > -\eta$;

(iii) $\lim (\log | a_n |)/\lambda_n = \lim (\log | A_n \exp [-\lambda_n \mathfrak{a}_F |])/\lambda_n$
$= \mathfrak{a}_F - \mathfrak{a}_F = 0, \lim (\log | b_n |)/\mu_n = 0;$

(iv) $\lim (\log n)/\lambda_n = \lim (\log n)/\mu_n = 0.$

Suppose now that there exists an infinite subsequence $\{\lambda_{n_i}\}$ of $\{\lambda_n\}$ such that $\lambda_{n_i} \neq \mu_m$ for all positive integers i and m. Let $\{\lambda_{n_i}\}$ be the maximum such subsequence of $\{\lambda_n\}$. Let $\{\lambda_{m_j}\}$ be the complement in $\{\lambda_n\}$ of $\{\lambda_{n_i}\}$: each λ_{m_j} belongs both to $\{\lambda_n\}$ and to $\{\mu_n\}$. Let $\{\mu_{n_j}\}$ be the subsequence of $\{\mu_n\}$ which is identical with $\{\lambda_{m_j}\}$, and let $\{\mu_{m_k}\}$ denote its complement in $\{\mu_n\}$: $\lambda_{m_j} = \mu_{n_j}$ $(j = 1, 2, \cdots)$.

Write $\gamma(s) = f(s) - \phi(s)$. Then $\gamma(s)$ is holomorphic and "M" in $\sigma > -\eta$. In Theorem 1, replace $\phi(s)$ by $\gamma(s)$, A by S_f, and B by S_γ, which is empty. Here $\sigma_1 = \sigma_2 = -\eta$, $-\eta \leq \sigma^\infty(A) \leq 0$, $\sigma^\infty(B) = -\eta$. Then $\sigma^\infty = \max [\sigma_1 + \sigma^\infty(B), \sigma_2 + \sigma^\infty(A)] \leq -\eta$. The set $[A, B]^\infty$ is empty, so that by Theorem 1 $H(s) = H(f, \gamma)$ is holomorphic and "M" in $\sigma > -\eta$. The function $\gamma(s)$ is given by:

$$\gamma(s) = \sum a_n e^{-\lambda_n s} - \sum b_n e^{-\mu_n s} = \sum a_{n_i} e^{-\lambda_{n_i} s} + \sum (a_{m_j} - b_{n_j}) e^{-\lambda_{m_j} s} - \sum b_{m_k} e^{-\mu_{m_k} s}.$$

Then

$$H(s) = H(f, \gamma) = \sum a_{n_i}^2 \, e^{-\lambda_{n_i} s} + \sum a_{m_j} (a_{m_j} - b_{n_j}) e^{-\lambda_{m_j} s}.$$

Under the hypotheses, it is easily verified that neither series has an infinite abscissa of absolute convergence. The function $H(s)$ satisfies the conditions on $F(s)$ of Lemma 1, with

$$f_1(s) = \sum a_{n_i}^2 \, e^{-\lambda_{n_i} s}, \qquad f_2(s) = \sum a_{m_j} (a_{m_j} - b_{n_j}) \, e^{-\lambda_{m_j} s}, \qquad N = 2.$$

Also,

$$\mathfrak{a}_{f_1} = \limsup \, (1/\lambda_{n_i}) \log | a_{n_i}^2 | = 2 \limsup \, (-1/\lambda_{n_i}) \log | a_{n_i} | = 0,$$

by hypothesis (iii). Hence

$$\mathfrak{a} = \max_{i=1,2,} \mathfrak{a}_{f_i} \geq 0,$$

so that $\mathfrak{A}_H \geq 0$, by Lemma 1. This contradicts the previous assertion that $H(s)$ is holomorphic in $\sigma > -\eta$, and completes the proof of the assertion that all of the terms of the sequence $\{\lambda_n\}$, except at most a finite number, belong also to the sequence $\{\mu_n\}$.

Upon interchanging $F(s)$ and $\Phi(s)$ in the discussion above, it becomes apparent that the two sequences $\{\lambda_n\}$ and $\{\mu_n\}$ are identical, with the exception of at most a finite number of terms of each sequence.

Hypothesis (iii) of Theorem 4 is certainly essential. For any series satisfying a condition of type (iv) and having a sufficiently small abscissa of absolute convergence might be added to $F(s)$, creating a new function such that all of the other hypotheses would be satisfied without the conclusion being valid.

As part of his thesis [3], Mandelbrojt found sufficient conditions in order that a Taylor series have at least two singular points on its circle of convergence. His method of approach there suggested the ideas developed in the next few paragraphs.

Throughout the rest of this section, two points will be said to belong to the same "class" (modulo $2\pi i$) if each may be obtained from the other by the addition of an integral multiple of $2\pi i$. Each class of points has a representative in any horizontal, half-open strip of width 2π: that point of the strip from which each point of the class may be obtained by the addition of a suitable integral multiple of $2\pi i$. In this way a many-one correspondence may be set up between points of any set in the plane, and points of a "representative" set in a given, horizontal, half-open strip of width 2π.

Consider the set of those singular points of a given function $f(s)$ which lie in the half-plane $\sigma > \sigma_1$. Let $R_f(\sigma > \sigma_1)$ denote the corresponding representative set (modulo $2\pi i$) in the strip $0 \leq t < 2\pi$. To each point of $R_f(\sigma > \sigma_1)$ corresponds one or more singular point of $f(s)$ with the same real part, whose imaginary parts differ from its imaginary part by integral multiples of 2π.

Let $R_f(\sigma = \sigma_0)$ denote the set representative of all points of S_f with real part σ_0. It consists of a point or points on the line segment $\sigma = \sigma_0$, $0 \leq t < 2\pi$.

This notation is useful in stating the following theorem, giving sufficient conditions in order that $R_f(\sigma = \mathfrak{a}_F)$ contain at least two points. The following definition will also be required.

Definition. The poles of a function in a region are said to have *finite maximum order* if there exists a positive integer K such that the order of every pole in the region is less than or equal to K. The smallest such integer is the maximum order, M, of the poles.

Theorem 5. *Suppose that* $F(s) = \Sigma A_n e^{-\lambda_n s}$ *satisfies the following conditions:*

(i) $F(s)$ *is "M" in* $\sigma > \mathfrak{a}_F - \eta$ *for some positive number* η;

(ii) $F(s)$ *has a singular point,* s_0, *on* $\sigma = \mathfrak{a}_F$; *also, all singular points* *of* $F(s)$ *in* $\sigma > \mathfrak{A}_F - \eta$ *lie on* $\sigma = \mathfrak{A}_F$, *those at points* $s_0 + 2\pi ki$ (*k an integer*) *being poles of finite maximum order* M;

(iii) *there exists a subsequence* $\{\lambda_n^{(0)}\}$ *of* $\{\lambda_n\}$ *such that* $\{\lambda_n^{(0)}\} \neq \lambda_m + k$ *for all* m *and* n, *and for all positive integers* k *less than or equal to* M; *and such that* $\limsup (\log |A_n^{(0)}|)/\lambda_n^{(0)} = \mathfrak{A}_F$, *where* $A_n^{(0)}$ *is the coefficient of the* $\exp\{-\lambda_n^{(0)}s\}$ *term in the series* $\Sigma A_n e^{-\lambda_n s}$;

(iv) $\lim (\log n)/\lambda_n = 0$.

Then $R_F(\sigma = \mathfrak{A}_F)$ *contains at least two points.*

Proof. Write

$$f(s) = F(s + s_0) = \sum A_n e^{-\lambda_n(s+s_0)} = \sum a_n e^{-\lambda_n s},$$

where $a_n = A_n e^{-\lambda_n s_0}$. Then $\mathfrak{A}_F = 0$, and $f(s)$ satisfies the following conditions:

(i) $f(s)$ is "M" in $\sigma > -\eta$;

(ii) the point $s = 0$ is singular for $f(s)$; also, all singular points of $f(s)$ in $\sigma > -\eta$ lie on $\sigma = 0$, those at points $2\pi ki$ (k an integer) being poles of finite maximum order M;

(iii) $$\limsup (\log |a_n^{(0)}|)/\lambda_n^{(0)} = \limsup (1/\lambda_n^{(0)}) \log \{|A_n^{(0)}| \exp [-\lambda_n^{(0)} \mathfrak{A}_F]\} = 0.$$

It is desired to prove that $R_f(\sigma = 0)$ contains at least two points. Suppose that this is not true. Then the only singular points of $f(s)$ in $\sigma > -\eta$ are poles of finite maximum order M at certain points $2\pi ki$. Write

$$f_1(s) = f(s)(1 - e^{-s})^M = f(s) \sum_{j=0}^{M} (-1)^j C_{j,M} e^{-js}$$
$$= \sum_{n=0}^{\infty} a_n e^{-\lambda_n s} \sum_{j=0}^{M} (-1)^j C_{j,M} e^{-js} = \sum_{j=0}^{M} (-1)^j C_{j,M} \sum_{n=0}^{\infty} a_n e^{-(\lambda_n + j)s},$$

where $C_{j,M}$ is the binomial coefficient.

The function $f_1(s)$ may thus be expressed as the sum of a finite number of functions, each of which is evidently "M" in $\sigma > -\eta$, being the product of an "M" function and an entire function bounded in $\sigma > -\eta$. Moreover, $f_1(s)$ is holomorphic in $\sigma > -\eta$, since the zeros of $(1 - e^{-s})^M$ "cancel" the poles of $f(s)$. By Lemma 2 and the remark following, $f_1(s)$ is "M" in $\sigma > -\eta$.

Denote by $\{\lambda_n^{(j,k,\cdots,p)}\}$ the sequence of λ_n's (if any), each of which belongs to the sequences $\{\lambda_n\}$, $\{\lambda_n + j\}$, $\{\lambda_n + k\}, \cdots$, $\{\lambda_n + p\}$, and to no others:

$$\{\lambda_n^{(j,k,\cdots,p)}\} = \{\lambda_n\} \cdot \{\lambda_n + j\} \cdot \{\lambda_n + k\} \cdots \{\lambda_n + p\};$$

$j, k, \cdots, p$ are positive integers less than or equal to M. For example, $\{\lambda_n^{(0)}\}$ denotes the sequence of positive numbers which belong to $\{\lambda_n\}$ but to no $\{\lambda_n + k\}$ $(k = 1, 2, \cdots, M)$; $\{\lambda_n^{(2)}\}$ is the sequence of λ_n's each of which belongs only to $\{\lambda_n\}$ and to $\{\lambda_n + 2\}$; $\{\lambda_n^{(1,2)}\}$ is the sequence of λ_n's, each of which belongs only to $\{\lambda_n\}$, $\{\lambda_n + 1\}$, and $\{\lambda_n + 2\}$. Designate by $a_n^{(j,k,\cdots,p)}$ the coefficient of $\exp\{-\lambda_n^{(j,k,\cdots,p)} s\}$.

The sum of series representing $f_1(s)$ may be rewritten as a Dirichlet series: $f_1(s) = \Sigma d_n e^{-\mu_n s}$. If $\{\mu_{n_i}\}$ is the sequence of all terms of $\{\mu_n\}$ which do not belong to $\{\lambda_n\}$, then $f_1(s)$ may be written:

$$\begin{aligned} f_1(s) = {} & \sum d_{n_i} e^{-\mu_{n_i} s} + \sum a_n^{(0)} e^{-\lambda_n^{(0)} s} + (1 - C_{1,M}) \sum a_n^{(1)} e^{-\lambda_n^{(0)} s} \\ & + (1 + C_{2,M}) \sum a_n^{(2)} e^{-\lambda_n^{(2)} s} + \cdots + [1 + (-1)^M C_{M,M}] \sum a_n^{(M)} e^{-\lambda_n^{(M)} s} \\ & + (1 - C_{1,M} + C_{2,M}) \sum a_n^{(1,2)} e^{-\lambda_n^{(1,2)} s} + \cdots \end{aligned}$$

the sum of a finite number of series. Then

$$\begin{aligned} H(s) = H(f, f_1) = {} & \sum [a_n^{(0)}]^2 e^{-\lambda_n^{(0)} s} \\ & + [1 - C_{1,M}] \sum [a_n^{(1)}]^2 e^{-\lambda_n^{(1)} s} + \cdots + \\ & + [1 + (-1)^M C_{1,M}] \sum [a_n^{(M)}]^2 e^{-\lambda_n^{(M)} s} \\ & + (1 - C_{1,M} + C_{2,M}) \sum [a_n^{(1,2)}]^2 e^{-\lambda_n^{(1,2)} s} + \cdots, \end{aligned} \tag{9}$$

the sum of a finite number of series. In Theorem 1, replace $\phi(s)$ by $f_1(s)$, A by S_f, and B by S_{f_1}, which is empty. Then $\sigma_1 = \sigma_2 = -\eta$, $\sigma^\infty(A) \leq 0$, $\sigma^\infty(B) = -\eta$, so that

$$\sigma^\infty = \max [\sigma_1 + \sigma^\infty(B), \sigma_2 + \sigma^\infty(A)] \leq -\eta.$$

By Theorem 1, $H(s) = H(f, f_1)$ is both "M" and holomorphic in $\sigma > -\eta$; $[A, B]^\infty$ is certainly empty since B is empty.

But $H(s)$ may be written as the sum of a finite number of functions as in (9). If

$$\gamma(s) = \sum [a_n^{(0)}]^2 e^{-\lambda_n^{(0)} s},$$

then, by hypotheses (iii) and (iv),

$$\begin{aligned} a_\gamma &= \limsup (1/\lambda_n^{(0)}) \log | a_n^{(0)} |^2 \\ &= 2 \limsup (1/\lambda_n^{(0)}) \log | a_n^{(0)} | = 0. \end{aligned}$$

Thus at least one of the functions of which $H(s)$ is the sum has a zero abscissa of absolute convergence; by Lemma 1, $\mathfrak{A}_H = 0$. But this contradicts the previous assertion that $H(s)$ is holomorphic in $\sigma > -\eta$, and completes the proof of the theorem.

The purpose of supposing as hypothesis (iv) that $\lim (\log n)/\lambda_n = 0$ rather than that $\lim (\log n)/\lambda_n^{(0)} = 0$, was to assure that none of the functions of which $H(s)$ is a sum should have a positively infinite abscissa of absolute convergence.

The hypothesis $\limsup (\log | A_n^{(0)} |)/\lambda_n^{(0)} = \mathfrak{a}_F$ is certainly not impossible of fulfillment, in view of the fact that $\mathfrak{a}_F = \limsup (\log | A_n |)/\lambda_n$. It is essential, as the following example shows. Write $\phi(s) = \Sigma e^{-ns} = 1/(1 - e^{-s})$. Define $g(s) = \Sigma c_n e^{-\mu_n s}$ so that $\lim (\log n)/\mu_n = 0$, $\mathfrak{a}_g < -\eta$, no μ_n is an integer, $\mu_n \neq \mu_m + 1$ for all n, m (for example, $\mu_n = n + 1/n$, $c_n = e^{-n}$). Write $F(s) = \phi(s) + g(s) = \Sigma A_n e^{-\lambda_n s}$; then $\mathfrak{a}_F = 0$. Hypotheses (i), (ii), and (iv) are satisfied. So also is the first part of (iii), the subsequence $\{\lambda_n^{(0)}\}$ being the sequence $\{\mu_n\}$. But

$$\limsup (1/\lambda_n^{(0)}) \log | A_n^{(0)} | = \limsup (\log | c_n |)/\mu_n \leq -\eta \neq 0;$$

and the conclusion is not valid. In fact, $R_F(\sigma = 0)$ contains only one point, $s = 0$. The theorem therefore fails if the hypothesis

$$\limsup (\log | A_n^{(0)} |)/\lambda_n^{(0)} = \mathfrak{a}_F$$

is not included.

Theorem 6, which follows, is an almost immediate consequence of Theorem 5.

THEOREM 6. *Suppose that* $F(s) = \Sigma A_n e^{-\lambda_n s}$ *is "M" in* $\sigma > \mathfrak{a}_F - \eta$ *for some positive* η, *and has as its only singular points in* $\sigma > \mathfrak{a}_F - \eta$ *poles of finite maximum order at certain points* $s_0 + 2\pi k i$ *for some point* s_0 *on* $\sigma = \mathfrak{a}_F$; *suppose further that* $\lim (\log n)/\lambda_n = 0$, *and that*

$$\lim (\log | a_n |)/\lambda_n = \mathfrak{a}_F.$$

Then there exists a positive integer n_0 *such that*

$$F(s) = \sum_{q=0}^{n_0} e^{-\lambda_q s} T_q(s);$$

where $T_q(s)$ *is a finite or infinite Taylor-Dirichlet series.*

Proof. All of the hypotheses of Theorem 5 are satisfied except (iii), and the conclusion does not hold. Also $\lim (\log | a_n |)/\lambda_n = \mathfrak{a}_F$, so that $\limsup (\log | A_n^{(0)} |)/\lambda_n = \mathfrak{a}_F$ is satisfied for any infinite subsequence $\{\lambda_n^{(0)}\}$. It follows that there exists no infinite subsequence $\{\lambda_n^{(0)}\}$ of $\{\lambda_n\}$ such that $\lambda_n^{(0)} \neq \lambda_m + k$ for all m and n, and for all positive integers k less than or equal to M. There then exists a positive integer n_0, such that for $n > n_0$, to each λ_n there corresponds a λ_{n_1} less than λ_n which differs from λ_n by a positive integer less than M. That is, to each λ_n $(n > n_0)$ there corresponds a λ_{n_1} and a positive integer R_{n_1} such that $0 < R_{n_1} \leq M$, and $\lambda_n = \lambda_{n_1} + R_{n_1}$. If $n_1 > n_0$, there is an n_2 such that $\lambda_{n_1} = \lambda_{n_2} + R_{n_2}$, where R_{n_2} is a positive integer, and $0 < R_{n_2} \leq M$. Evidently an n_k less than or equal to n_0 will be reached after repeating this a finite number of times, such that $\lambda_n = \lambda_{n_k} + (R_{n_1} + R_{n_2} + \cdots + R_{n_k})$. Thus to every n greater than n_0 there corresponds a positive integer $p(n) = n_k$, less than or equal to n_0, and a positive integer $R(n) = R_{n_1} + R_{n_2} + \cdots + R_{n_k}$ such that $\lambda_n = \lambda_{p(n)} + R(n)$. Hence to each positive integer q, less than or equal to n_0, there corresponds an empty, a finite, or an infinite sequence of λ_n's, each term of which is obtained from the preceding by adding a positive integer; the members of these sequences exhaust the sequence $\{\lambda_n\}$. That is, the sequence $\{\lambda_n\}$ is composed of a finite number n_0 of finite or infinite (or empty) sequences $\{\lambda_q + n_i^{(q)}\}$, where $\{n_i^{(q)}\}$ is, for each q less than or equal to n_0, a finite or infinite sequence of positive integers. Therefore $F(s)$ may be written:

$$F(s) = \sum_{q=0}^{n_0} e^{-\lambda_q s} \sum_{i=0}^{r_q} d_i^{(q)} e^{-n_i^{(q)} s},$$

where $0 \leq r_q \leq \infty$, and where the $d_i^{(q)}$ are complex numbers. On writing

$$T_q(s) = \sum_{i=0}^{r_q} d_i^{(q)} e^{-n_i^{(q)} s},$$

the form of the statement of the theorem is obtained.

3. Examples. In view of the fact that in most of the theorems of this paper the functions concerned are supposed "M" in some half-plane, it is pertinent to ask whether such functions exist. Obviously all periodic functions with an imaginary period have this property; in particular, those given by Taylor-Dirichlet series, which have period $2\pi i$. However, the class of functions which are "M" in open half-planes containing their half-planes of absolute convergence is far wider than that of periodic functions. For an example of a wide class of such functions, let $\phi(s)$ be given by a Taylor-

Dirichlet series $\Sigma b_n e^{-ns}$ having a finite abscissa of absolute convergence $\mathfrak{a}_\Phi$. Let $f(s)$ be any function given by a Dirichlet series $\Sigma a_n e^{-\lambda_n s}$ such that $\mathfrak{a}_f < \mathfrak{a}_\Phi$. If $\mathfrak{a}_f < \sigma_1 < \mathfrak{a}_\Phi$, and if $\psi(s) = f(s) + \phi(s)$, then $\psi(s)$ is not only "M," but actually bounded except for singularities, in $\sigma > \sigma_1$. For certainly $\phi(s)$, being represented by a Taylor-Dirichlet series, is bounded except for singularities in $\sigma > \sigma_1$, and $f(s)$ is bounded in $\sigma > \sigma_1$, since $\sigma_1 > \mathfrak{a}_f$. The function $\psi(s)$ thus furnishes the desired example. A much wider class of functions bounded except for singularities is given by Mandelbrojt ([4], page 4).

It is worth noting, in the example given above, that if $\psi(s) = \Sigma c_n e^{-\nu_n s}$, then $\lim\, (\log |c_n|)/\nu_n$ does not necessarily exist. Such an "M" function could not satisfy the conditions of Theorems 4 and 6. In fact, the sequence $(\log |c_n|)/\nu_n$ has at least two distinct limit points, if $\lim\, (\log n)/\lambda_n = 0$; for then $\mathfrak{a}_f = \limsup\, (\log |a_n|)/\lambda_n$ is one limit point of the sequence $\{(\log |c_n|)/\nu_n\}$, and $\mathfrak{a}_\Phi = \limsup\, (\log |b_n|)/n$ is another. To assure the non-triviality of Theorems 4 and 6, it will, therefore, be well to have an example of a non-periodic function $\Sigma c_n e^{-\nu_n s}$ which is "M" in an open half-plane containing in its interior its half-plane of absolute convergence, and for which $\lim\, (\log |c_n|)/\nu_n$ exists. An "M" function of the form $\Sigma e^{-\nu_n s}$ would satisfy this condition, for $\lim\, (\log 1)/\nu_n = 0$. In constructing such a function, let $\mu_1, \mu_2, \cdots, \mu_n$ ($M \geq 2$) be a finite set of linearly independent positive numbers, so that $x_1\mu_1 + x_2\mu_2 + \cdots + x_M\mu_m = 0$, when $x_1, \cdots, x_M$ are integers, only if $x_1 = x_2 = \cdots = x_M = 0$. Then

$$x_1\mu_1 + x_2\mu_2 + \cdots + x_M\mu_m = y_1\mu_1 + \cdots + y_M\mu_m$$

for x_i and y_i $(i = 1, 2, \cdots, M)$ integers, only if $x_i = y_i$ $(i = 1, 2, \cdots, M)$. Write

$$f_j(s) = \sum e^{-\mu_j s} = \frac{1}{1 - e^{-\mu_j s}}, \qquad (j = 1, 2, \cdots, M).$$

Each $f_j(s)$ is periodic, with period $2\pi i/\mu_j$. Also, no $f_j(s)$ has a finite zero. Hence

$$F(s) = \prod_{j=1}^{M} f_j(s)$$

is singular at each finite point which is singular for some $f_j(s)$, and at no others. It follows that for every real number σ_1, $F(s)$ is bounded except for singularities in $\sigma > \sigma_1$, and $F(s)$ is "M" in the entire plane. But clearly $F(s) = \Sigma e^{-\nu_n s}$, where the set of non-negative numbers $\{\nu_n\}$ is the set of all numbers of the form $x_1\mu_1 + x_2\mu_2 + \cdots + x_M\mu_M$, where each x_i $(i = 1, 2, \cdots, M)$ takes on all non-negative integral values. The function $F(s)$ is evidently not periodic, but is quasi-periodic in the sense of Bohl and Esclangon (cf. H. Bohr, "Zur Theorie der fast periodischen Funktionen," *Acta Math.*, vol. 46 (1925), p. 114, Theorem III).

THE RICE INSTITUTE.

BIBLIOGRAPHY

1. Bernstein, V., *Séries de Dirichlet*, Gauthier-Villars, Paris, 1933.
2. Brunk, H. D., "Theorems of Composition for Dirichlet Series," *Duke Mathematical Journal*, vol. 12 (1945), pp. 1-21.
3. Mandelbrojt, S., "Sur les séries de Taylor qui présentant des lacunes," *Annales de l'École Normale*, vol. 40 (1923), p. 428.
4. Mandelbrojt, S., "Contribution à la théorie du prolongement analytique des séries de Dirichlet," *Acta Math.*, vol. 55 (1930), pp. 1-32.
5. Mandelbrojt, S., "Sur la recherche des points singuliers d'une série de Dirichlet," *Bulletin de la Société Mathématique de France*, vol. 57 (1929), pp. 78-103.
6. Mandelbrojt, S., *Dirichlet Series*, The Rice Institute Pamphlet, 1944.
7. Valiron, G., "Théorie générale des séries de Dirichlet," *Mémorial des Sciences Mathématiques*, Gauthiers-Villars et C^{ie}, Paris, 1928.

EXTENSIONS OF PLANE CONTINUA MAPPINGS.*

By G. T. Whyburn.

1. Introduction. Extension theorems for continuous transformations of the sort which extend the range of definition and retain continuity are well known—(see, for example, Chapter 1 of [1]). However, our knowledge of the extensibility of mappings so as to preserve given essential properties, such as monotoneity or interiority, in addition to continuity is much more limited. Some steps in this direction have been taken by the present author before (see Chapters 9 and 11 of [2]); and in this paper a further study of the problem is made. In particular, we shall deal largely with the situation in which a continuous transformation of a given sort acts on a locally connected continuum imbedded in a sphere (or a plane) and shall seek to extend the action of the mapping to the whole sphere without sacrificing its given vital properties. Indeed, in some cases it will be found possible to effect the extension so that the extended mapping is of a much stronger and more restricted type than was the original one.

All spaces used are assumed to be separable and metric. Frequently the space is of a much more restricted character and its nature is specified in the separate results.

If $A \subset X$, $B \subset Y$ and $f(A) = B$ and $g(X) = Y$ are continuous, g is said to be an extension of f to X provided

$$g(x) = f(x)$$

for all $x \in A$.

If $A \subset X$, $f(A) = B$ and $g(X) = Y$ are continuous and $B' = g(A)$, g is said to be an extension of f provided the transformations

$$f(A) = B \text{ and } g(A) = B'$$

are topologically equivalent, i. e., provided there exists a homeomorphism

$$h(B) = B'$$

such that

$$g(x) = hf(x), \qquad x \in A.$$

In other words, g is an extension of f provided we can imbed B topologically in Y so that

---

* Received April 25, 1945.

$$g(x) = f(x) \text{ for all } x \in A.$$

A continuous transformation $f(A) = B$ is said to be (i) *monotone,* if for each $y \in B$, $f^{-1}(y)$ is a continuum (i. e., compact and connected), (ii) *non-alternating* if for $x, y \in B$, $f^{-1}(y)$ separates no two points of $f^{-1}(x)$ in A, (iii) *interior* (or *open*) if the image of each open set in A is open in B, and (iv) *light* if for each $y \in B$, $f^{-1}(y)$ is totally disconnected.

A locally connected continuum is called (i) a *cactoid* (see [3]) if each of its true cyclic elements is a topological sphere, (ii) a *boundary curve* if each of its true cyclic elements is a simple closed curve, and (iii) a *dendrite* if it has no true cyclic elements.

For definitions of other terms used the reader is referred to the author's book *Analytic Topology* (see [2]).

2. Composite mappings on boundary curves.

It is not true in general that non-alternating mappings applied successively yield a non-alternating composite transformation, although this is true of monotone and of interior transformations. However, for light non-alternating transformations acting on boundary curves this sort of "group property" was shown in [2], Chapter 9, to hold. In this section, we proceed to prove the validity of this conclusion without the assumption of lightness.

(2.1) LEMMA. *Let $f(A) = B$ be non-alternating where A (hence also B) is a boundary curve. Let E_b be any true cyclic element of B and let $x, y \in E_b$ be non cut points of B such that if $X = f^{-1}(x)$, $Y = f^{-1}(y)$, then $X + Y$ contains no cut point of A. Then $A - (X + Y)$ consists of just two components, namely, $f^{-1}(R)$ and $f^{-1}(S)$, where R and S are the two components of $B - (x + y)$.*

Proof. Let E_a be the true cyclic element of A such that $f(E_a) \supset E_b$. Then since neither X nor Y can separate A and neither contains a cut point of A, it results that X and Y are each continua and furthermore $X + Y \subset E_a$. Also it follows from this that $A - (X + Y)$ has just two components, say R' and S', each containing an open arc of E_a with one end point in X and the other in Y. Since $B - (x + y) = R + S$, it results that one of these, say R', maps onto R and the other, S', maps onto S.

(2.2) THEOREM. *If A is a boundary curve and the mappings $f_1(A) = B$ and $f_2(B) = C$ are non-alternating, so also is the mapping $f(A) = f_2 f_1(A) = C$.*

Proof. Since the non-alternating image of a boundary curve is a boundary curve, it results from our hypothesis that both B and C are boundary curves. Let $x, y \in C$, $f^{-1}(x) = X$, $f^{-1}(y) = Y$. We must show that X does not separate Y in A. Let $X' = f_2^{-1}(x)$, $Y' = f_2^{-1}(y)$ and let W be the least A-set in B containing X'.

Now if $W \cdot Y' = 0$, let Q be a component of $B - W$ intersecting Y' and let $p = F(Q)$. Then since p non $\in Y'$, p cannot separate two points of Y' in B and hence $Y' \subset Q$. Thus since f_1 is non-alternating, $f_1^{-1}(Q)$ is a component of $A - f_1^{-1}(p)$ containing $f_1^{-1}(Y') = Y$ but containing no point of X. Thus X does not separate Y in this case.

If $W \cdot Y' \neq 0$, there exists a true cyclic element E_b of B intersecting both X' and Y'. Since f_2 is non-alternating, there exist disjoint arcs U and V of E_b containing $X' \cdot E_b$ and $Y' \cdot E_b$ respectively. Since the number of points of E_b which are either cut points of B or images under f_1 of cut points of A is countable, there exist points $a, b \in E_b$ which do not belong to this set and such that a and b separate U and V on E_b. Then $B - (a + b)$ consists of just two components R and S containing U and V respectively. Since X' can contain no cut point of B on $V \cdot E_b$ and Y' can contain no cut point of B on $U \cdot E_b$, we must have

$$X' \subset R, \qquad Y' \subset S.$$

Now by the lemma, $f^{-1}(R)$ and $f^{-1}(S)$ are connected and, since they contain $X = f_1^{-1}(X')$ and $Y = f^{-1}(Y')$, respectively, it results that X cannot separate Y in A.

3. Reduction to complementary domain closures.

Since the general problem under study amounts to the question of how to define a mapping in the complementary domains of the continuum A on a plane or sphere which will be identical with the given mapping on A and will retain its essential properties on the whole sphere or plane, it is of first importance to know to what extent these properties automatically hold on the complementary domain boundaries of A when they hold on A. In this section this question is treated briefly for monotone and for non-alternating transformations. For interior transformations this question is treated in 7 along with the general extensibility problem.

(3.1) THEOREM. *Let $f(M) = A$ be non-alternating, where M is a plane locally connected continuum and A is a dendrite. Then if B is the boundary of any complementary domain of M, f is non-alternating on B.*

Proof. Let C be any true cyclic element of B, let $x, y \in f(C) = D$ and let z be an interior point of the arc xy of A which is of order 2 in A. Then since $A - z$ consists of two components one containing x and the other y and f is non-alternating, $M - f^{-1}(z)$ consists of two distinct components R and Q containing $X = f^{-1}(x)$ and $Y = f^{-1}(y)$ respectively.

Now if f alternated on C, for some $x, y \in f(C)$ there would exist $x_1, x_2 \in X \cdot C$ and $y_1, y_2 \in Y \cdot C$ such that x_1 and x_2 separate y_1 and y_2 on C. But since R contains an arc $x'_1x'_2$ spanning C where $x'_1 \in y_1x_1y_2$, $x'_2 \in y_1x_2y_2$ and this arc must intersect Q, we have the contradiction that $R \cdot Q \neq 0$. Hence the mapping $f(C) = D$ is non-alternating. Thus f is non-alternating on each true cyclic element of B.

Further, if $y \in f(B)$ and $a, b \in B \cdot f^{-1}(y)$, the set $E(a, b)$ of all points separating a and b in B is identical with the set of all points separating a and b in M. Thus since f is non-alternating on M, $E(a, b) \subset f^{-1}(y)$. Hence by [VIII, 2.4] p. 139 of [2], f is non-alternating on B.

(3.2) *Example.* As an interesting case where (3.1) applies, let us consider the mapping of the lemniscate L: $\rho^2 = \sin 2\theta$ onto the dendrite D composed of the intervals of the x and y-axes joining $(-1, 0)$ to $(1, 0)$ and $(0, -1)$ to $(0, 1)$ obtained as follows: map the points given by $\theta = 0$, $\theta = \pi/4$ and $\theta = 5\pi/4$ into the origin O. Then map the points given by $\theta = \pi/8$, $\theta = 3\pi/8$, $\theta = 9\pi/8$ and $\theta = 11\pi/8$ into the end points $p_1 = (1, 0)$, $p_2 = (0, 1)$, $p_3 = (-1, 0)$ and $p_4 = (0, -1)$ respectively of D. Finally map the remaining 8 segments of L by pairs topologically into the four open segments of $D - O - p_1 - p_2 - p_3 - p_4$ so as to get a continuous mapping f of L onto D. It is then readily observed that f is non-alternating on L and also on each loop of L as required by our theorem.

That (3.1) does not hold if no restriction is imposed on the image set A is seen by the following example. Let M consist of a circle C together with two perpendicular diameters x_1px_2 and y_1py_2 of C. Let f map M by identifying x_1 and x_2 and also y_1 and y_2 but otherwise being topological. Then clearly f is non-alternating on M but not on C, the boundary of the unbounded complementary domain of M.

However, in case f is monotone our same conclusion follows without any limitation on A, as will now be demonstrated.

(3.3) THEOREM. *Any monotone transformation on a plane (or spherical) locally connected continuum M is non-alternating on the boundary of each complementary domain of M.*

Proof. Let B be the boundary of any such complementary domain of M, let C be any true cyclic element of B, let M' be the cyclic element of M containing C and let $f(C) = E \subset f(M) = A$. Now M' lies either in C plus its interior or in C plus its exterior, say in C plus its interior. Then if $x, y \epsilon E$ and $X = M' \cdot f^{-1}(x)$, $Y = M' \cdot f^{-1}(y)$, X and Y are continua lying in C plus its interior. Let α be a minimum arc of C containing X. If $\alpha \cdot Y = 0$, then $Y \cdot C$ separates no two points of $X \cdot C$ in C. If $\alpha \cdot Y \neq 0$ then $Y \cdot C$ must lie wholly in α, since $X \cdot Y = 0$; in this case let β be the minimum subarc of α containing Y. Then $\beta \cdot X = 0$ since $X \cdot Y = 0$. Thus $C - \beta$ is a connected subset of $C - Y \cdot C$ containing $X \cdot C$; so that again $Y \cdot C$ separates no two points of $X \cdot C$ in C.

Thus f is non-alternating on each true cyclic element C of B. Further, for any $y \epsilon A$ and any $a, b \epsilon f^{-1}(y) \cdot B$, we must have $E(a, b) \subset f^{-1}(y)$ since f is non-alternating on M and any point separating a and b in B also separates them in M. Therefore, just as in the proof of (3.1), it follows that f is non-alternating on B.

4. Extensions from complementary domain closures.

Once our definition of a given transformation on a plane or spherical continuum A is extended to the separate complementary domains of A so as to obtain the desired properties on the closures of these separate domains, the question arises as to what extent do these properties hold for the extended mapping on the whole sphere or plane. This section is devoted to the development of the needed results of this type for monotone and non-alternating transformations.

(4.1) THEOREM. *Any non-alternating transformation of a locally connected unicoherent continuum onto a dendrite is necessarily monotone.*

Proof. Let $f(A) = B$ be such a transformation. Let $y \epsilon B$, $Y = f^{-1}(y)$ and suppose that Y is not connected. Then if $Y = Y_1 + Y_2$ is a separation of Y, there exists a closed subset X of $A - Y$ which separates $y_1 \epsilon Y_1$ and $y_2 \epsilon Y_2$ in A; and since A is unicoherent, we can suppose that X is a continuum. Now since $f(X)$ and y are disjoint subcontinua of the dendrite B, there exists a point z of B which separates y and $f(X)$ in B. But then $f^{-1}(z)$ separates y_1 and y_2 in A. For any arc y_1y_2 in A must intersect X and thus must intersect $f^{-1}(z)$, since $f^{-1}(z)$ separates Y and X in A.

Definition. A mapping $f(A) = B$ on a compact set A will be said to be monotone on a subset A_0 of A provided $A_0 \cdot f^{-1}(y)$ is either empty or connected for each $y \epsilon B$.

(4.2) THEOREM. *Let M be a locally connected continuum on a sphere*

S and let $f(S) = D$ be a continuous mapping of S onto a dendrite D. Then if f is non-alternating on M and for each complementary domain R of M with boundary B, f is monotone on $\bar{R}$ and $f(\bar{R}) = f(B)$, then f is monotone on S.

Proof. By virtue of (4.1) we have only to show that f is non-alternating on S, because S is unicoherent. To this end, let $x, y \epsilon D$ and let $X = f^{-1}(x)$, $Y = f^{-1}(y)$. Since f is non-alternating on M, there exists a component Q of $M - Y \cdot M$ containing $X \cdot M$. Let K be any component of X and let R be any complementary domain of M intersecting X. Then by hypothesis if $B = F(R)$, $X \cdot \bar{R}$ is a continuum and $f(\bar{R}) = f(B)$. Hence there exists a point $x_0 \epsilon X \cdot B \subset X \cdot M$; and since $x_0 \epsilon X \cdot \bar{R}$, we have $X \cdot \bar{R} \cdot Q \neq 0$. Since this holds for all complementary domains R of M which intersect X it follows that X lies in the component of $S - Y$ containing Q. Accordingly Y separates no two points of X on S; and hence f is non-alternating and thus monotone on S.

(4.3) THEOREM. *Let A be a locally connected continuum on a sphere S. Any monotone mapping $f(S) = K$ of S onto a dendrite K which is monotone on each complementary domain of A and on its closure is non-alternating on A.*

Proof. Let $f(A) = B$ and $x, y \epsilon B$, $X = f^{-1}(x)$, $Y = f^{-1}(y)$. Then X is a continuum. Now for each complementary domain R of A with boundary F, since[1] f is non-alternating on F, there exists a component Q of $F - Y \cdot F$ containing $X \cdot F$. Hence if in each such case we replace $X \cdot R$ by Q in the set X we obtain a connected subset D of $A - A \cdot Y$ which contains $A \cdot X$. Accordingly it follows that the mapping $f(A) = B$ is non-alternating.

5. Monotone extensions of non-alternating transformations.

It has been shown previously by the author (see p. 180 of [2]) that any non-alternating transformation f on a circle C can be extended monotonically to the interior I of C, i.e., so that the resulting mapping is monotone on $C + I$. Further, this extension can be effected in an extremely simple manner by decomposing $C + I$ into the convex closures $f^{-1}(y)_c$ of the sets $f^{-1}(y)$, for $y \epsilon f(C)$, together with individual points of I not in any one of these sets $f^{-1}(y)_c$. Since other monotone extensions of f to I would exist, we shall refer to this particular one as the *monotone extension of f to I given by convexification of inverses.* This extension will serve as a starting point for most of the extensions to be developed in this paper. In general this extension involves an enlargement of the image space. However, in most cases which will concern us, the image space remains unchanged as indicated in the following theorem.

---

[1] This is a consequence of an unpublished result of the author's to the effect that a monotone mapping of the closure of a plane simply connected region R having property S onto a dendrite which is monotone on R is non-alternating on $F(R)$. See a forthcoming note, "Boundary alternation of monotone mappings," in the *Duke Mathematical Journal.*

(5.1) THEOREM. *Let* $f(C) = D$ *be non-alternating where* C *is a circle and* D *is a dendrite and let* ϕ *be the monotone extension of* f *to the interior* I *of* C *given by convexification of inverses. Then* $\phi(C + I) = D$.

Proof. Let $E = I + C$. Then since ϕ is monotone, $\phi(E)$ is a hemicactoid with boundary curve D. Thus D is the base set of $\phi(E)$. Now if $\phi(E) - D \neq 0$, there would exist a point $y \epsilon D$ and a component R of $\phi(E) - y$ containing no point of D. But then $\phi^{-1}(R)$ would be a component of $E - \phi^{-1}(y)$ not intersecting C. Clearly this is impossible since $\phi^{-1}(y)$ is convex.

We have the following interesting corollary of a geometric nature:

(5.11) COROLLARY. *If* $f(C) = D$ *is any non-alternating mapping of the circle* C *onto a dendrite* D, *any point within* C *is contained uniquely in the convex closure of* $f^{-1}(y)$ *for some* $y \epsilon B$.

(5.2) LEMMA. *If* A *is a locally connected continuum which is the boundary of a plane (or spherical) bounded region* R *and* $f(A) = D$ *is any non-alternating mapping of* A *onto a dendrite* D, *there exists a monotone extension* ϕ *of* f *to* $R + A$ *such that* $\phi(R + A) = D$ *and which is monotone also on* R.

Proof. By [VIII, (9.8)], p. 161, of [2] there exists a non-alternating light mapping g of the circle C plus its interior I onto $A + R$, such that $g(C) = A$, $g(I) = R$ and the mapping $g(C) = A$ is non-alternating and the mapping $g(I) = R$ is topological. By (2.2) the mapping $fg(C) = D$ is non-alternating. Let h be the monotone extension of the mapping fg to I given by convexification of inverses. Since D is a dendrite, we have $h(C + I) = D$ by (5.1).

Now for each $x \epsilon R + A$, let us define

$$\phi(x) = hg^{-1}(x).$$

Then if $x \epsilon A$, we have

$$\phi(x) = fgg^{-1}(x) = f(x),$$

so that ϕ is an extension of f. Further for $y \epsilon D$, $\phi^{-1}(y) = gh^{-1}(y)$; and since

$h^{-1}(y)$ is a continuum, it results that ϕ is monotone. Clearly $\phi(R+A) = h(C+I) = D$; and ϕ is monotone on R, since $I \cdot h^{-1}(y)$ is connected for each $y \epsilon D$ and the mapping $g(I) = R$ is topological.

Since (5.2) effects the extension of f to the region R by reducing it back to the case where the boundary is a circle and then uses the "convexification of inverses" method, we shall speak of the extension ϕ obtained in (5.1) also as *the monotone extension given by convexification of inverses.*

(5.3) THEOREM. *Any non-alternating mapping $f(A) = D$ of a locally connected continuum A on a sphere S onto a dendrite D is monotonically extensible to S—and indeed so that the extended mapping is monotone on each complementary domain and on its closure and each such domain closure has the same image as the boundary of the domain.*

Proof. Let R be any complementary domain of A on S and let B be its boundary. By (3.1), f is non-alternating on B; and since $f(B)$ is a subdendrite of D, by (5.2) there exists a monotone extension ϕ of f to R so that $\phi(\bar{R}) = f(B)$. Let ϕ be defined on S as the mapping obtained by extending f to each such R in this way. Then clearly ϕ is monotone on each R and on the closure of each such R and $\phi(\bar{R}) = \phi(B)$. Further, since f is continuous, since $\phi(x) \equiv f(x)$ for $x \epsilon A$ and since the complementary domains of A on S form a null sequence at most, it results that the mapping $\phi(S) = D$ is continuous. Finally, ϕ is non-alternating on A since it is identical with f on A; and therefore, by (4.2), ϕ is monotone on S. This completes the proof.

Definition. Since other monotone extensions of f to S would exist, we shall refer to the particular one set up in the preceding theorem and its proof as the *monotone extension given by the convexification of inverses method.*

Results (4.3) and (5.3) yield at once

(5.4) THEOREM. *In order that the continuous mapping $f(A) = D$ of a locally connected continuum A on a sphere S onto a dendrite D be non-alternating it is necessary and sufficient that f be monotonically extensible to S so that the extended mapping ϕ is monotone on each complementary domain of A and on its closure and so that $\phi(S) = D$.*

Thus if A is a locally connected continuum on a sphere S, the class $[f]$ of all non-alternating mappings of A into dendrites D is equivalent to the class $[\phi]$ of all monotone mappings on S which map S into dendrites and which are monotone on the complementary domains of A and on their closures. For each such $f \epsilon [f]$ is associated uniquely with a $\phi \epsilon [\phi]$ such that $f(x) \equiv \phi(x)$ for $x \epsilon A$.

Remark. The reader may find it of interest to follow through the steps of the extension as it would be effected by the procedure described in (5.2) and (5.3) in the case of the example of the lemniscate mapping into a cross as defined in (3.2).

6. Monotone extensions of monotone transformations. Since any monotone transformation is non-alternating, it results from the theorems in **5** that any monotone transformation of a locally connected continuum on a sphere onto a dendrite is monotonically extensible to the whole sphere and, indeed, so that the extended mapping satisfies the extra conditions in (5.3). However, for monotone transformations the problem can be handled in a much simpler way.

For let A be any compact subset of a separable metric space M. Then *any monotone mapping* $f(A) = B$ *on* A *can be extended monotonically to* M. To do this it is only necessary to decompose M into the sets $f^{-1}(y)$, $y \in B$, together with the individual points of $M - A$; let M' be the hyperspace of this decomposition and let $\phi(M) = M'$ be the associated transformation. Then ϕ is monotone and, since ϕf^{-1} is a homeomorphism h mapping B onto a subset B' of M' so that $hf(x) \equiv \phi f^{-1} f(x) = \phi(x)$ for $x \in A$, ϕ is an extension of f to M.

Since any monotone image of a sphere is a cactoid, we have

(6.1) THEOREM. *Any monotone image of a compact set on a sphere is topologically contained in a cactoid.*

For locally connected continua A on a sphere S it is interesting to note that the method of convexification of inverses affords a second—and in general different—method of effecting monotone extensions of arbitrary monotone transformations on A, whether the image of A is a dendrite or not. For if f is monotone on A it is non-alternating on the boundary of each complementary domain of A by (3.3). Accordingly the method of convexification of inverses affords a continuous extension ϕ of f to S as described in the proof of (5.3), except that it is no longer assured that each complementary domain closure will have the same image as the boundary of the domain. However, for any $y \in \phi(S)$, $\phi^{-1}(y)$ either intersects A or else lies entirely in a single complementary domain of A; and since f is monotone and ϕ is monotone on each complementary domain closure, it follows in either case that ϕ is monotone.

It may be noted that when the former (more general) extension method is used, A would be an inverse set under the extended mapping, whereas when the method of convexification of inverses is used this is not usually the case. Both methods in general involve enlargement of the image space. It would be of interest to compare the two methods of extension more fully.

Although we have been able, in the case of monotone mappings, to dispense with the restriction that the image set be a dendrite, this cannot be done in the case of non-alternating transformations in general. Further, no significant assertion can be made concerning the structure of the non-alternating image of a locally connected continuum on a plane or sphere. These facts are exhibited by

(6.2) EXAMPLE. *There exists a non-alternating light mapping of a 2-cell, or of a regular curve A on a sphere S, onto a set containing the Q_ω which is topological on the boundary of each complementary domain of A and yet which cannot be extended monotonically to S.*

For let A be either a hemisphere on S or the sum of two Sierpinski regular curves [4] on S having a common side I_0 but otherwise being disjoint. In the former case let I denote a non-degenerate arc lying entirely within A and in the latter let I be a subinterval of I_0 containing neither end point of I_0. In either case let $f(I) = Q_\omega$ be any continuous light mapping. Decompose A into the sets $f^{-1}(y)$, $y \in Q_\omega$, together with the individual points of $A - I$ and let B be the hyperspace and $g(A) = B$ the associated transformation. Then B contains $gf^{-1}(Q_\omega)$ which is homeomorphic with Q_ω. Further, no set $g^{-1}(z)$, $z \in B$, separates A so that g is non-alternating.

Clearly in either case g is topological on the boundary of each complementary domain of A since no such boundary intersects I in more than one point. In neither case can there exist a monotone extension of g to S. For let $y \in g(A)$ such that $g^{-1}(y)$ is non-degenerate. Then $g^{-1}(y)$ is a totally disconnected subset of I and there exists no continuum in $S - A + g^{-1}(y)$ containing two distinct points of $g^{-1}(y)$.

7. Monotone extensions of interior transformations. In view of results already obtained the question naturally arises as to whether the monotone extension we have obtained for a given non-alternating mapping of a locally connected continuum on a sphere will be interior provided the original mapping is interior. This section is devoted to developing conditions under which this will be true.

(7.1) THEOREM. *Let $f(A) = B$ be interior and non-alternating where A is a plane (or spherical) locally connected continuum. Then for any cut point p of B, $f^{-1}(p)$ intersects the boundary F of any complementary domain D of A in at most two points which are accessible from a given component Q of $A - f^{-1}(p)$.*

Proof. For if three points x_1, x_2, x_3 of $F \cdot f^{-1}(p)$ were accessible from Q, $D + Q + x_1 + x_2 + x_3$ would contain a θ—curve $\theta = ax_1b + ax_2b + ax_3b$, where $a \in D$, $b \in Q$. But there exists a second component R of $A - f^{-1}(p)$; and by the non-alternating property, $R = f^{-1}f(R)$. Hence by interiority, $F(R) = f^{-1}(p)$. Thus $F(R) \supset x_1 + x_2 + x_3$ and $R \cdot \theta = 0$; and clearly this is impossible.

By an exactly similar argument we can prove

(7.2) THEOREM. *Let $f(A) = B$ be interior and non-alternating where A is a plane (or spherical) locally connected continuum. Then for any point p of B which cuts B into at least three components, at most two points of $f^{-1}(p)$ are accessible from each of two components of $A - f^{-1}(p)$.*

Since if a region has property S all its boundary points are accessible, we have the following corollary [assuming the same situation as in (7.1) or (7.2)].

(7.21) COROLLARY. *If for each $p \in B$, each component of $A - f^{-1}(p)$ has property S, then*

(a) *for any cut point p of B, $f^{-1}(p)$ intersects the boundary of each complementary domain of A in at most two points, and*

(b) *for any point p which cuts B into at least three components, $f^{-1}(p)$ consists of at most two points.*

Now if f is light, the condition in (7.21) is automatically satisfied. Hence we have

(7.22) COROLLARY. *If $f(A) = B$ is interior, non-alternating and light, where A is a locally connected continuum on a plane or a sphere, conclusions* (a) *and* (b) *or* (7.21) *hold.*

As a standing hypothesis for statements (1)-(11) which follow, let $f(A) = B$ be interior, light and non-alternating where A is a locally connected continuum on a sphere S and B is a dendrite.

(1) For any end point y of B, $f^{-1}(y)$ is a single point.

(2) For any branch point y of B, $f^{-1}(y)$ consists of at most two points.

(3) For any $y \in B$, $f^{-1}(y)$ intersects the boundary of each complementary domain of A in at most two points.

Proof. By (1) and (2) we need only consider the case where y cuts B into exactly two components. Hence $f^{-1}(y)$ cuts A into just two components

R and Q and every point of $f^{-1}(y)$ is accessible from both R and Q. Accordingly at most two points of $f^{-1}(y)$ could lie on the boundary of any given complementary domain of A.

(4) If C is a simple closed curve in the boundary of any complementary domain R of A, then $f(C)$ is a simple arc ab and $C \cdot f^{-1}(a)$ and $C \cdot f^{-1}(b)$ are single points, whereas $C \cdot f^{-1}(x)$, $x \epsilon ab - a - b$, is a pair of points. Thus f is interior on C.

(5) If A is cyclic, f maps the boundary of each complementary domain of A interiorly and $(2-1)$ onto an arc of B.

(6) If no pair of points disconnects A into more than two components, B is necessarily a simple arc.

Proof. If B had a branch point y, by (2) $f^{-1}(y)$ would be a pair of points in A cutting A into at least three components.

Let $\phi(S) = B$ be the monotone extension of f to S, obtained by extending f monotonically in the complementary domain boundaries of A by *the convexification of inverse sets* process.

(7) For each end point y of B, $\phi^{-1}(y) = f^{-1}(y) =$ a single point.

(8) For each point y of B of order 2, $\phi^{-1}(y)$ is a simple closed curve.

Proof. Let R and Q be the two components of $A - f^{-1}(y)$. Since ϕ is monotone, $\phi^{-1}(y)$ is a continuum separating R and Q on S. Further, since the part of $\phi^{-1}(y)$ in any complementary domain of A is an open simple arc and $f^{-1}(y)$ is totally disconnected, $\phi^{-1}(y)$ is locally connected. Also since any complementary domain of A containing an arc $ab - (a+b)$ of $\phi^{-1}(y)$ is divided into two regions which are separated by $\phi^{-1}(y)$, it follows that $\phi^{-1}(y)$ separates R and Q irreducibly on S. Accordingly $\phi^{-1}(y)$ is a simple closed curve.

(9) If the image of each complementary domain boundary of A is a free arc in B, then ϕ is interior.

Proof. For if $x \epsilon A$, and U is any neighborhood of x, $f(U \cdot A)$ has $f(x)$ as an interior point relative to B. Hence also $\phi(U)$ has $\phi(x) = f(x)$ as an interior point relative to B, since $\phi(U) \supset f(U \cdot A)$.

Also, if $x \epsilon R$, a complementary domain of A, then since $R \cdot \phi^{-1}\phi(x)$ is an open arc dividing R into two regions R_1 and R_2 mapping under ϕ onto opposite sides of $\phi(x)$ in the arc $f[F(R)]$ and each having x on its boundary, it results that if U is any neighborhood of x, $\phi(U)$ has $\phi(x)$ as an interior point relative to B.

(10) If B is a simple arc, ϕ is interior.

(11) If the image of each complementary domain boundary of A is an arc, then f is interior on each complementary domain boundary.

Proof. It is non-alternating there and hence is interior, since any non-alternating light transformation onto an arc is interior.

(7.3) THEOREM. *Any light interior and non-alternating transformation* $f(A) = ab$ *of a locally connected continuum* A *on a sphere* S *onto a simple arc* ab *can be extended to a monotone interior mapping* $\phi(S) = ab$. *Further,* $\phi^{-1}(a)$ *and* $\phi^{-1}(b)$ *are single points while* $\phi^{-1}(y)$ *is a simple closed curve for any* $y \in ab - a - b$.

Now since any 1-dimensional locally connected continuum A which is identical with its cyclic chain $C(a, b)$, $a, b \in A$, can be mapped [5] onto an interval by a light non-alternating interior transformation, (7.3) yields the following result.

(7.31) EXISTENCE THEOREM. *If* A *is any* 1-*dimensional locally connected continuum on the sphere* S *which is identical with the cyclic chain* $C(a, b)$ *in* A, *there exists a monotone interior mapping* ϕ *of the whole sphere onto the interval* $[0, 1]$ *so that* ϕ *is light, interior and non-alternating on* A *and* $\phi^{-1}(0) = a$, $\phi^{-1}(1) = b$ *and for* $0 < y < b$, $\phi^{-1}(y)$ *is a simple closed curve separating* a *and* b.

The assumption of lightness is essential for the validity of (7.3) as is readily seen by a simple example. The extra condition needed in the absence of any lightness assumption is exhibited in the following theorem concerning monotone interior extensibility of the general non-alternating interior mappings in the situation under consideration.

(7.4) THEOREM. *An interior non-alternating mapping* $f(A) = ab$ *of a locally connected continuum* A *on a sphere* S *onto a non-degenerate simple arc* ab *can be extended to a monotone interior mapping* $g(S) = ab$ *of the whole sphere onto* ab *if and only if each component of* $S - X$ *intersects* A, *where* X *is any component of a set* $f^{-1}(y)$, $y \in ab$.

Proof. To establish the sufficiency of the condition, let $f(A) = B$ be interior and non-alternating and let

$$f = f_2 f_1, \quad f_1(A) = A' \text{ monotone}, \quad f_2(A') = ab \text{ light},$$

be the monotone—light factorization of f. Let

$$g_1(S) = K, \qquad K \supset A'$$

be the monotone extension of f_1 to S given by decomposing S into the sets $f^{-1}(y)$, $y \in A'$, and single points of $S - A$ and considering A' as imbedded in the hyperspace K of this decomposition (see **6**). Then K is a cactoid. Further, since $f^{-1}(a)$ and $f^{-1}(b)$ are connected, $a' = g_1 f^{-1}(a)$ and $b' = g_1 f^{-1}(b)$ are single points; and it results at once from our hypothesis that K is identical with the cyclic chain $C(a', b')$ in K.

For suppose there is a component R of $K - C(a', b')$. Then if $p = F(R)$, $g_1^{-1}(p)$ is a component X of $f^{-1}fg_1^{-1}(p)$; and since g_1 is monotone, $g_1^{-1}(R)$ contains a point x of A. But this means that $g_1^{-1}(R)$ contains either $f^{-1}(a)$ or $f^{-1}(b)$, say $f^{-1}(a)$, because every component of $A - f^{-1}fg_1^{-1}(p)$ contains either $f^{-1}(a)$ or $f^{-1}(b)$. Clearly this is impossible since then R would contain $g_1 f^{-1}(a) = a'$ which is absurd.

Now consider the mapping $f_2(A') = ab$. This mapping is non-alternating, interior and light. Further, if C_i is an arbitrary true cyclic element of $K = C(a', b')$, f_2 is non-alternating, light and interior on $A' \cdot C_i$.

Thus since the image $f_2(A' \cdot C_i)$ is a subarc $x_i y_i$ of ab, the mapping f admits of a monotone interior extension to all of C_i as given in (7.3). Let this extension be effected in each true cyclic element C_i of K and let $g_2(K) = ab$ be the resulting extension of f_2 to K. Since $E(a', b') \subset A'$ and since the arcs $x_i y_i$ can overlap only at their endpoints, clearly g_2 is single valued, continuous and monotone. Further, since $E(a', b')$ maps topologically[2] under f_2, it results at once that it also maps topologically under g_2 and thus that g_2 is interior, since g_2 is interior on each C_i.

Finally, let the decomposition of S into the sets $[g_1^{-1}g_2^{-1}(y)]$, $y \in ab$, generate the mapping $g(S) = ab$. Then g is monotone, since both g_1 and g_2 are monotone. Also g is identical with f on A, because $f = f_2 f_1$ and g_1 and g_2 are extensions of f_1 and f_2 respectively. Finally, g is interior. For since $g(A) = ab = g(S)$, g is interior at any point of A, and if x is in a component R of $S - A$, g_1 is topological and g_2 is interior on R so that g is interior at x, since $g = g_2 g_1$. This completes the sufficiency proof.

To prove the necessity, suppose that the mapping $f(A) = ab$ admits a monotone interior extension $g(S) = ab$ to the whole sphere S but that for some component X of $f^{-1}(y)$ for some $y \in ab$, there is a component R of $S - X$ with $S \cdot A = 0$. Since g is interior, $g(\bar{R}) \neq y$. Let $z \in g(\bar{R}) - y$. Then $g^{-1}(z) \cdot R \neq 0$ since $g(X) = y$; and $g^{-1}(z) \cdot (A - X) \neq 0$ since $g(A) \supset z$. But this is impossible since $g^{-1}(z)$ is connected and does not intersect X. This completes the proof.

---

[2] This statement is easily verified. Also it is a consequence of a much more general result concerning non-alternating interior mappings. See the author's paper, "On uniqueness of the inverse of a transformation," *Duke Mathematical Journal*, vol. 12 (1945), pp. 317-323.

It should be noted that in case the given mapping f is light, the condition in (7.4) is automatically satisfied. Hence (7.4) includes (7.3) as a special case.

The following examples shed light on questions which arise in connection with extensibility of interior transformations on various curves lying on a sphere or on a plane.

(7.51) EXAMPLE. *There exists a light interior mapping of a θ-curve onto an interval which is not interior on the boundary of one of the complementary domains to the curve.*

For let A be a circle C together with the horizontal diameter ao_2 of C. Let the upper semicircle of C be lettered $ao_1b_1o_2$ and the lower one $ao_3b_2o_2$ by choosing points in the indicated orders. Let $a'o'b'$ be an interval. Then map A into $a'o'b'$ by sending a into a', o_1, o_2 and o_3 into o', b_1 and b_2 into b' and sending the residual open intervals of A topologically into either $a'o' - (a' + o')$ or $o'b' - (o' + b')$ so as to obtain a continuous mapping f of A into $a'o'b'$. Then f is light and interior but is not interior on C.

(7.52) EXAMPLE. *There exists a light interior non-alternating mapping of a θ-curve into a dendrite (indeed a T-curve) T which admits no monotone interior extension to the whole sphere so that the image is still T.*

For let A consist of a circle C together with the horizontal diameter o_1yo_2 where y is the center of C. Let xo_1z and xo_2z be the right and left semicircles respectively of C and let R be the upper region into which o_1yo_2 divides the interior of C. Let f map A into the part T of A to the right of and on the line xyz by reflection of the part of C to the left in this line.

Then f is interior, light and non-alternating but admits no extension ϕ of the sort described above. For if so then since R is connected and has x and y on its boundary, its image under ϕ must contain o_1. Hence there is a point u in R with $\phi(u) = o_1$. By interiority of ϕ there is a point v in R with $\phi(v) = v' \in o_1z - o_1$. But $\phi^{-1}(v')$ then intersects both R and o_1zo_2 but does not intersect the boundary $o_1xo_2 + o_1yo_2$ of R, contrary to the fact that it must be connected.

8. Conclusion. It may be remarked that nearly all our extensions in this paper have been effected by the "convexification of inverses" method and in only a few cases is the image space augmented. Other means of securing extensions could surely be applied and would doubtless yield interesting and significant conclusions even for the types of transformations considered here. A study of the general extension question from this viewpoint is under way and will probably be dealt with in a later paper.

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

[1]. Alexandroff-Hopf, *Topologie*, J. Springer, Berlin, 1935.

[2]. G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications, New York, 1942.

[3]. R. L. Moore, "Concerning upper semi-continuous collections," *Monatshefte für Mathematik und Physik*, vol. 36 (1929), pp. 81-88.

[4]. W. Sierpinski, *Prace Matematyczno Fizyczne*, vol. 27 (1915); or Knaster and Kuratowski, *Bulletin of the American Mathematical Society*, vol. 33 (1927), pp. 106-107.

[5]. G. T. Whyburn, "The existence of certain transformations," *Duke Mathematical Journal*, vol. 5 (1938), pp. 647-655.

A THEORY OF CRITICAL POINTS.*

By Magnus R. Hestenes.

1. Introduction. Let $f(x) = f(x_1, \cdots, x_n)$ be a function defined on a neighborhood of a closed set S in euclidean n-space. Suppose that the points of S at which $f_{x_i} = 0$ are interior to S and that the determinant $| f_{x_i x_k} |$ is different from zero at these points. A point at which $f_{x_i} = 0$ is called a critical point and its index is defined to be the negative index of the quadratic form $f_{x_i x_k}\xi_i\xi_k$ at the point. Let m_k be the number of critical points on S of index k and let r_k be the k-th connectivity of S. Then the numbers $h_k = m_k - r_k$ satisfy the relations

$$h_k \geqq h_{k-1} - h_{k-2} + \cdots + (-1)^{k-1}h_0 \geqq 0 \tag{1.1}$$

if the outer normal derivative of $f(x)$ on the boundary of S is positive. This result, which is due to Morse (III),[1] has been generalized by Morse (XVII) to functions on a metric space and, in particular, to functions of curves appearing in the simple problems of the calculus of variations. The special cases where $k \leqq 1$ were treated earlier by G. D. Birkhoff (II) and Poincaré (I).

In the earlier papers the relations (1.1) were obtained by observing the variation of the connectivities of the sets $f \leqq c$ as c increased through critical values. More recently, a second method of approach has been used. This method is based essentially on a maximum-minimum principle and seems to have its beginning with G. D. Birkhoff in 1927 in his book on Dynamical Systems (V, p. 135). Here Birkhoff establishes the existence of closed extremals of index $k > 0$ with the help of a concept which, in the terminology of Morse, is the inferior limit of a k-cycle. Birkhoff treated only a special case. Later Morse (X, p 272) showed that the relations $m_k \geqq r_k$ could be obtained by the use of inferior limits of absolute k-cycles. Birkhoff and Hestenes (XI) then pointed out that the complete relations (1.1) could be obtained if one used the inferior limits of modular k-cycles as well as those of absolute k-cycles. They did not use the term "inferior limit" but used an equivalent concept.

---

* Received June 8, 1940; Revised January 25, 1942. Presented to the American Mathematical Society April 11, 1936, under the title "Minimax principle in the Calculus of variations"; April 9, 1938, under the title "A theory of critical points of functions in the calculus of variations."

[1] Roman numerals in parentheses refer to the list of papers at the end of this paper. Further references on the theory of critical points are found in these papers.

Moreover deformations were used where homologies should have been used. Later Morse (XII, XVII) introduced the notion of superior limit of an absolute k-cycle. One readily sees that the superior limit of an absolute k-cycle, if finite, is in general the inferior limit of a modular $(k+1)$-cycle. By the use of superior and inferior limits of cycles Morse has developed an elaborate theory of critical points of functions on a metric space. An interesting book on the theories of Morse has been written recently by Seifert and Threlfall (XVIII). References to the various writers on this subject are found in the papers of Morse.

In the present paper a theory of critical points is developed on a metric space which is quite different from that given by Morse but is equivalent to it. The method used is basically the one outlined by Birkhoff and Hestenes (XI, XIV), modified in many ways. It consists essentially of a study of the properties of k-cycles u on F mod F', where F and F' are restricted to belong to a given class $\mathfrak{F}$ with prescribed properties. The properties assigned to $\mathfrak{F}$ are ones enjoyed by the sets determined by inequalities of the form $f(x) \leqq c$, where $f(x)$ is a real valued function and c is a constant. It is shown that the topological theory here developed is based only on these properties. In fact even critical points can be defined in a topological way without specifying the functions $f(x)$ used to determine the class $\mathfrak{F}$, at hand. This development is made in Sections **2-11** below. It is pointed out in Section **12** that our assumptions are not the most general ones that could be used and an indication is given how this extension can be carried out. The extension is not developed in detail in the present paper because in the first place the theory here developed is sufficiently general for our purposes and in the second place it is simpler not to do so.

The paper is divided into two parts; the first is concerned with the topological aspects of the theory and the second with an application to the calculus of variations. Two topological definitions of critical points are given, the first by means of homologies (**5**) and the second by means of homotopic deformations (**7**) in a manner similar to that given by Morse. Analogues of the critical point relations are given in **10** and **11** and applications to functions of n real variables are given in **13** and **14**.

I. TOPOLOGICAL CONSIDERATIONS.

2. Vietoris chains on a metric space *S*. The topological part of the present paper is based on the concept of Vietoris chains. It is the purpose of the present section to define Vietoris chains and to recall the properties of these chains that will be useful in the present paper. In general, proofs of these properties will be omitted. The definition of Vietoris cycles here given is somewhat different from the one usually given but is equivalent to it. The author has found this definition to be a very convenient one. It emphasizes a property of k-cycles that is perhaps the most essential property, namely, that in an homology relation any sequence of vertex k-chains defining a Vietoris k-cycle can be replaced by a subsequence.

Vietoris chains are defined in terms of algebraic or vertex chains and these in turn are defined in terms of simplices. By an *oriented k-simplex z on a metric space S* will be meant a class composed of an ordered set P of $k+1$ points on S and all ordered sets obtained from P by an even permutation of its points. If two or more points of P are coincident, the simplex will be said to be *degenerate.* To every set of $k+1$ points $(k>0)$ on S there correspond two simplices, one of which will be taken to be positively oriented and the other negatively oriented. They will be denoted respectively by the symbols z and $-z$. If the simplex is degenerate, z and $-z$ are identical. A single point on S will be said to define a positively oriented 0-simplex.

Let Λ be an additive abelian group of elements λ. *Algebraic* or *vertex k-chains* over Λ are defined in the usual manner as the formal sum

$$u = \Sigma\lambda_i z_i = \lambda_1 z_1 + \cdots + \lambda_m z_m, \tag{2.1}$$

where $\lambda_1, \cdots, \lambda_m$ are elements in Λ and the z's are positively oriented k-simplices, not necessarily distinct. We identify with u the k-chain obtained from u by replacing the sum $\lambda_i z_i + \lambda_j z_j$ by $(\lambda_i + \lambda_j) z_i$ when $z_i = z_j$ and the k-chain obtained from u by deleting a term $\lambda_i z_i$ in which $\lambda_i = 0$ or z_i is degenerate. A vertex k-chain therefore can be represented uniquely in the form (2.1) in which the simplices are non-degenerate, distinct and have non-zero coefficients. Such a form will be called the *reduced form* of u. Two k-chains are identical if they have the same reduced form. The vertices of the k-simplexes in the reduced form of u will be called *vertices* of u and the maximum of the diameters of the simplices in the reduced form will be called the *norm* of u. The boundary u' of u is defined in the usual manner. We have $(u')' = 0$. By the sum $u + v$ of the k-chain (2.1) and $v = \Sigma\lambda^*_j z^*_j$ will be meant the k-chain $\Sigma\lambda_i z_i + \Sigma\lambda^*_j z^*_j$. It is clear that $(u+v)' = u' + v'$.

An infinite sequence (u_n) of vertex k-chains, whose norms converge to zero as n becomes infinite, will be called a *Vietoris k-chain u on S* if the closure of the point set composed of the vertices of the vertex k-chains $u_1, u_2, \cdots$ forms a compact set U. The points in U will be called *vertices* of u. If U is null, we say that u is null and write $u = 0$. Let $u = (u_n)$, $v = (v_n)$ be two

Vietoris k-chains. The sequences $(u_n + v_n)$ and $(u_n - v_n)$ define Vietoris k-chains which we shall call the sum $u + v$ and the difference $u - v$ of u and v. We say that $u = v$ in case $u - v = 0$. By the *boundary* u' of a k-chain $u = (u_n)$ will be meant the $(k-1)$-chain defined by the sequence (u'_n), where u'_n is the boundary of u_n. We have the relations $(u')' = 0$, $(u + v)' = u' + v'$.

A k-chain u will be said to be *on a subset* F of S if its vertices are on F. The null k-chain is on every subset of S and, in particular, is on the null set. Let u be a k-chain on F and let F' be a subset of F such that u' is on F'. We shall say that the k-chain u is *homologous to zero* on F mod F', written $u \sim 0$ on F mod F', if there is a $(k+1)$-chain w on F such that the k-chain $w' - u$ is on F'. Two k-chains u and v on F will be said to be *homologous on F mod F'*, written $u \sim v$ on F mod F', if $u - v \sim 0$ on F mod F'. When $F = S$ the phrase "on F" will be omitted. Likewise when $F' = 0$ the phrase "mod F'" will be omitted.

By a *subchain* v of a k-chain $u = (u_n)$ will be meant a k-chain defined by an infinite subsequence (v_n) of (u_n). Consider now two sets F, F' on S such that $F \supset F'$. By a *k-cycle u on F mod F'* will be meant a k-chain u on F whose boundary u' is on F' and which is such that the relation, $u \sim v$ on F mod F', holds for every subchain v of u. It is easily seen that if u is a k-cycle on F mod F' and v is any k-chain satisfying the relation, $u \sim v$ on F mod F', then v is also a k-cycle on F mod F'. In particular, every subchain of a k-cycle on F mod F' is a k-cycle on F mod F'. The k-cycles here defined are Vietoris k-cycles, as can be seen by interpreting the result described in the following lemma in terms of e-homologies.

LEMMA 2.1. *Let $F \supset F'$ be subsets of S and let $u = (u_n)$ be a k-chain on F whose boundary u' is on F'. Then u is a k-cycle on F mod F' if and only if it is homologous on F mod F' to the subchain v of u defined by the sequence (u_{n+1}).*

It is clear from the definition of k-cycles that if u is a k-cycle on F mod F', then $u \sim v$ on F mod F', where $v = (u_{n+1})$. Conversely, suppose that this homology relation holds. Then there is a $(k+1)$-chain $w = (w_n)$ on F such that the k-chain $w' - u + v$ is on F', and hence such that the vertex k-chains $w'_n - u_n + v_n$ $(n = 1, 2, \cdots)$ are on F'. Consider now a subchain $\bar{v} = (u_{n_i})$ of $u = (u_n)$. Set $\bar{w}_i = 0$ if $n_i = i$ and set $\bar{w}_i = w_i + w_{i+1} + \cdots + w_{n_i - 1}$ otherwise. Then $\bar{w}_i - u_i + u_{n_i} = 0$ when $n_i = i$ and when $n_i > i$ the vertex k-chain

$$\bar{w}_i - u_i + u_{n_i} = \sum_{j=i}^{j=n_i-1} (w'_j - u_j + u_{j+1}),$$

is on F'. Setting $\bar{w} = (\bar{w}_n)$ it follows that the Vietoris k-chain $\bar{w}' - u + \bar{v}$ is on F' and hence that $u \sim \bar{v}$ on F mod F'. This completes the proof of Lemma 2. 1.

In order to insure that our k-cycles have the properties described in the next two lemmas, we make the following

ASSUMPTION. *Let F and $F' \subset F$ be compact sets and let k be a positive integer. Given a constant $e > 0$ there is a positive constant $e' < e$ such that if u is a vertex k-chain on F of norm at most e' with u' on F' and e'' is a positive constant, there exist on F a vertex k-chain v of norm at most e'' and a $(k+1)$-chain w of norm at most e, such that $w' - u + v$ and v' are on F'.*

This assumption holds for the cases in which we are interested. In particular it holds when Λ is a field.

The following lemma will be used frequently.

LEMMA 2. 2. *Let u be a k-cycle on F mod F' and let G be a subset of u such that $u \sim 0$ on F mod $F' + G$. There is a k-cycle v on $F' + G$ mod F' such that $u \sim v$ on F mod F'. In fact, if F' and G are closed, there is a k-cycle v on G mod $F'G$ such that $u \sim v$ on F mod F'.*

In the proof one can assume, without loss of generality, that the sets $F, F', F' + G$ are compact. The lemma can be established by an argument like that used by Morse (XVII, Theorem 3. 1, p. 396) for the case in which Λ is a field. In a similar manner (cf. Morse XVII, pp. 396, 392, Theorems 3. 2 and 3. 3 and Lemma 2. 2), one obtains the following:

LEMMA 2. 3. *Let $F \supset F'$ and G be closed subsets of S and let u be a k-cycle on F mod F':*

1. *If $G \supset F$ and $u \sim 0$ on G mod F', there is a $(k+1)$-cycle w on G mod F such that $u \sim w'$ on F mod F'.*

2. *If $G \subset F'$ and $u' \sim 0$ on F' mod G, where u' is the boundary of u, there is a k-cycle v on F mod G such that $u \sim v$ on F mod F'.*

3. *If F and F' are compact and $u \sim 0$ on N mod N' for every pair of neighborhoods N of F and N' of F', then $u \sim 0$ on F mod F'.*

3. Admissible classes $\mathfrak{F}$ of sets F on S.

Consider for the moment a real single valued lower semicontinuous function $f(x)$ on S and let $\mathfrak{F}$ be the class of all sets F on S determined by an inequality of the form $f(x) \leqq c$, where c is a constant or $\pm \infty$. The topological part of the theory of critical points of $f(x)$ on S here developed is essentially a theory of k-cycles on F mod F', where F and F' belong to $\mathfrak{F}$. It is based on the following properties of $\mathfrak{F}$: (1) *Given two sets in* $\mathfrak{F}$, *one is a subset of the other*; (2) *the intersection of the sets belonging to a subclass of* $\mathfrak{F}$ *is in* $\mathfrak{F}$; (3) *the sets in* $\mathfrak{F}$ *are closed*; (4) *the null set* $F = 0$ *and the set* $F = S$ *are in* $\mathfrak{F}$. In order to emphasize the dependence of our theory on these properties, we shall begin with an arbitrary class $\mathfrak{F}$ having these properties. Such a class will be called *admissible*. The class $\mathfrak{F}$ related to a function $f(x)$ as described above will be called the *admissible class determined by* $f(x)$. Since the results described below depend only on the four properties listed above, it follows that two functions $f(x)$ and $g(x)$, which determine the same admissible class $\mathfrak{F}$, give rise to the same topological theory. They have accordingly the same critical k-sets (**5**), the same homotopic critical points (**7**), etc. Moreover, they are connected by a relation of the form $f(x) = h[g(x)]$, where $h(t)$ is a non-decreasing function of a real variable t.

Consider now a particular admissible class $\mathfrak{F}$. In the sequel the symbol (u, F, F') will be used to denote a k-cycle u on $F \bmod F'$ for which F, F' are in $\mathfrak{F}$ and $F \supset F'$. Two k-cycles (u, F, F') and (v, G, G') will be said to be *homologous*, written $(u, F, F') \sim (v, G, G')$, if $u \sim v$ on $F + G \bmod F' + G'$. It is clear that either $F + G = F$ or $F + G = G$ in view of property (1) for the admissible class $\mathfrak{F}$. A k-cycle (u, F, F') will be said to be *homologous to zero*, written $(u, F, F') \sim 0$, if $u \sim 0$ on $F \bmod F'$.

We shall center our attention on special k-cycles called minimizing k-cycles. By a *minimizing k-cycle* (u, F, F') will be meant one such that the relation $F \subset G$ holds for every k-cycle (v, G, G') that is homologous to it. In view of this definition, the null k-cycle $(0, 0, 0)$ is minimizing. This definition is a simple modification of one used by Birkhoff and Hestenes (XI, XIV) in a similar situation. In the terminology of Morse (XVII, p. 408), the k-chain u belonging to a non-null minimizing k-cycle (u, F, F') would be called a k-cap.

LEMMA 3.1. *A k-cycle* (u, F, F') *is minimizing if and only if there is no proper subset* F'' *of* F *in* $\mathfrak{F}$ *such that* $u \sim 0$ *on* $F \bmod F' + F''$.

This criterion is obviously satisfied when $F = 0$. Suppose therefore that $F \neq 0$. If $(u, F, F') \sim 0$, the relation $u \sim 0$ on $F \bmod F' + F''$ holds with $F'' = 0$. In this case the k-cycle (u, F, F') is not minimizing. We can assume, therefore, that $(u, F, F') \not\sim 0$. Then $F' \neq F$. If there is a proper subset F'' of F in $\mathfrak{F}$ such that $u \sim 0$ on $F \bmod F' + F''$, then (u, F, F') is homologous to a k-cycle of the form $(v, F'', F'F'')$, by Lemma 2.2 with $G = F''$. Since F'' is a proper subset of F, the k-cycle (u, F, F') cannot be minimizing in this event. Finally, suppose that the k-cycle (u, F, F') is not minimizing and let (v, G, G') be a k-cycle homologous to it such that $FG \neq F$. Then G is a proper subset of F and the relation $u \sim 0$ on $F \bmod F' + G$ holds, since $u \sim v$ on $F \bmod F' + G'$. The homology relation described in the lemma accordingly holds with $F'' = G$. This completes the proof of the lemma.

COROLLARY. *If (u, F, F') is a minimizing k-cycle and G is a set in $\mathfrak{F}$ such that u is on G, then $F \subset G$.*

For since u is on F and on G it is also on FG. Hence $u \sim 0$ on $F \bmod F' + FG$. By the last lemma with $F'' = FG$, it follows that $FG = F$, that is, $F \subset G$, as was to be proved.

If $\mathfrak{F}$ is the class of sets related to a function $f(x)$, as described above, the corollary states that for a minimizing k-cycle (u, F, F') the least upper bounds of $f(x)$ on F and on the set of vertices of u are the same.

In order to facilitate generalizations in **12** below and to avoid excessive repetitions of the phrase "a proper subset of F in $\mathfrak{F}$" we shall define a proper subset of F in $\mathfrak{F}$ to be a *primary subset* of F. With the definition in mind we can prove the following:

LEMMA 3.2. *Let (u, F, F') be a non-null minimizing k-cycle. The set F' is a primary subset of F. If (v, G, G') is a k-cycle homologous to (u, F, F'), it is minimizing if and only if $F = G$ and G' is a primary subset of F. The k-cycle $(u, F, F' + F'')$ is minimizing for every primary subset F'' of F.*

The set F' is a primary subset of F, by virtue of the criterion described in the last lemma. Consider now a k-cycle (v, G, G') that is homologous to (u, F, F'). If it is minimizing, then obviously $F = G$ and G' is a primary subset of F. Conversely, if these relations hold, the k-cycle (v, G, G') must be minimizing. Otherwise there would exist, by Lemma 3.1, a proper subset G'' of G in $\mathfrak{F}$ such that $v \sim 0$ on $F \bmod G' + G''$. Moreover the set $F'' = G' + G''$ would be a proper subset of $F = G$, since G' is a primary subset of F. From the relation $u \sim v$ on $F \bmod F' + G'$ it would follow that the relation $u \sim 0$ on $F \bmod F' + F''$ would hold with $F'' = G' + G''$. This is impossible, by Lemma 3.1. Hence (v, G, G') is minimizing in this case. The last statement in the lemma follows from the preceding one and the fact that the sum of two primary subsets of F is a primary subset of F.

LEMMA 3.3. *Two minimizing k-cycles (u, F, F') and (v, G, G'), not both null, are homologous to a third minimizing k-cycle if and only if $F = G$ and there is a primary subset F'' of F such that $u \sim v$ on $F \bmod F''$.*

For let (w, H, H') be a minimizing k-cycle that is homologous to two minimizing k-cycles (u, F, F') and (v, G, G'). Then $F = H = G$ and $u \sim v$ on $F \bmod F''$ with $F'' = F' + G' + H'$. If $F \neq 0$, the set F'' is a primary subset of F, by the last lemma. Conversely if $u \sim v$ on $F \bmod F''$, where F'' is a primary subset of F, then (u, F, F') and (v, G, G') are homologous to the minimizing k-cycle $(u, F, F' + G' + F'')$. This proves the lemma.

4. Existence of minimizing k-cycles.

As a first existence theorem we have

THEOREM 4.1. *If the sets $F \neq S$ in $\mathfrak{F}$ are compact, every k-cycle (u, F, F') is homologous to a minimizing k-cycle (v, G, G') having $G' \subset F'$. In fact the k-cycle (v, G, G') can be chosen so that the $(k-1)$-cycle $(v', G', 0)$ is minimizing, where v' is the boundary of v.*

In order to prove the first statement in the theorem, consider a k-cycle (u, F, F') having $F \neq S$. If $(u, F, F') \sim 0$, choose $(v, G, G') = (0, 0, 0)$. Suppose, therefore, that $(u, F, F') \not\sim 0$. Let $\mathfrak{F}''$ be the class of all subsets F'' of F in $\mathfrak{F}$ such that the relation $u \sim 0$ on $F \bmod F' + F''$ holds. Let G be the intersection of all sets in $\mathfrak{F}''$. Then G is in $\mathfrak{F}$, by property (2) of $\mathfrak{F}$. In fact G is in $\mathfrak{F}''$. To prove this result, observe that on every neighborhood N' of G there is a set F'' belonging to $\mathfrak{F}''$. Consequently $u \sim 0$ on $F \bmod F' + N'$ for every neighborhood N' of G. By the use of part 3 of Lemma 2.3 it is seen that $u \sim 0$ on $F \bmod F' + G$ and hence that G is in $\mathfrak{F}''$. From Lemma 2.2 it follows that there is a k-cycle of the form $(v, G, F'G)$ that is homologous to (u, F, F'). We shall show that $(v, G, F'G)$ is minimizing. If it were not so there would exist, by Lemma 3.1, a proper subset G'' of G in $\mathfrak{F}$ such that $v \sim 0$ on $G \bmod F'G + G''$. Hence $u \sim 0$ on $F \bmod F' + G''$ since (u, F, F') and $(v, G, F'G)$ are homologous. But this would imply that G'' is in $\mathfrak{F}''$ contrary to our choice of G as the intersection of the sets in $\mathfrak{F}''$. The k-cycle $(v, G, F'G)$ is accordingly minimizing and the first statement in the theorem is proved for the case in which $F \neq S$. Suppose now that $F = S$ and that (u, F, F') is not minimizing. Then there is a k-cycle (w, H, H') with $H \neq F$ and $H' \subset F'$ that is homologous to (u, F, F'). We have just seen that there is a minimizing k-cycle (v, G, G') homologous to (w, H, H') such that $G' \subset H'$. It is clear that $(u, F, F') \sim (v, G, G')$. This completes the proof of the first part of the theorem.

In order to prove the last statement in the theorem select a minimizing k-cycle (v, G, G') homologous to (u, F, F') with $G' \subset F'$. If the $(k-1)$-cycle $(v', G', 0)$ is not minimizing, there is a proper subset F'' of G' in $\mathfrak{F}$ such that $v' \sim 0$ on G' mod F''. Let $\mathfrak{F}''$ be the class of all sets F'' of this type and let G'' be the intersection of the sets in $\mathfrak{F}''$. Then, as in the last paragraph, $v' \sim 0$ on G' mod G'', since G' and G'' are compact. By virtue of part 2 of Lemma 2. 3 there is a k-cycle (w, G, G'') that is homologous to (v, G, G') and hence to (u, F, F'). From our choice of the set G'' it follows that the $(k-1)$-cycle $(w', G'', 0)$ is minimizing. The k-cycle (w, G, G''), being homologous to (v, G, G'), is minimizing. This completes the proof of Theorem 4. 1.

THEOREM 4. 2. *Suppose that the sets* $F \neq S$ *in* $\mathfrak{F}$ *are compact, and let* (u, F, F') *be a non-null minimizing* k*-cycle such that* $u \sim 0$ *on* S *mod* F'. *There exists a minimizing* $(k+1)$*-cycle* (v, G, F) *such that* $(v', F, 0) \sim (u, F, F')$ *where* v' *is the boundary of* v. *In fact* (v, G, F) *can be chosen so that, given a proper subset* G'' *of* G *in* $\mathfrak{F}$, *one has* $u \not\sim 0$ *on* G'' *mod* F'.

For let (u, F, F') be a minimizing k-cycle such that $u \sim 0$ on S mod F' and let $\mathfrak{F}''$ be the maximal subclass of $\mathfrak{F}$ such that $u \sim 0$ on F'' mod F' for every set F'' in $\mathfrak{F}''$. Then the intersection G of the sets in $\mathfrak{F}''$ has the property that $u \sim 0$ on G mod F', by virtue of part 3 of Lemma 2. 3 and the compactness of the sets $F \neq S$ in $\mathfrak{F}$. Using part 2 of the same lemma one finds that there is a $(k+1)$-cycle (v, G, F) such that $u \sim v'$ on F mod F', that is, such that $(u, F, F') \sim (v', F, 0)$. This proves the theorem.

THEOREM 4. 3. *The conclusions described in Theorems* 4. 1 *and* 4. 2 *hold true if* $\mathfrak{F}$ *has the property that given a set* $F \neq S$ *in* $\mathfrak{F}$ *and a compact subset* C *of* F *there is a compact subset* G *of* F *in* $\mathfrak{F}$ *such that* $C \subset G$.

The proof is like that given above. One uses the additional fact that if u is a k-cycle on F mod F' there exist compact subsets G, G' of F, F' respectively such that u is a k-cycle on G mod G'.

The hypotheses in Theorem 4. 3 are satisfied, for example, when the class $\mathfrak{F}$ is the class of all sets determined by inequalities of the form $f(x) \leqq c$ or $f(x) < c$, where $f(x)$ is continuous on S, provided the sets of the form $f(x) \leqq c$ are compact. In this case the assumption of closure of the sets in F is not satisfied. This assumption is not needed in the proof of Theorem 4. 3.

5. Critical k-sets. In the introduction critical points of index k were defined for a function of n real variables. It is the purpose of the present section to define analogues of these points by means of homologies. The analogue of a critical point of index k here given is, in general, not a single point but a set of points, which we shall call a critical k-set. It will be seen in **14** below that a critical point of index k forms a critical k-set.

Consider now a minimizing k-cycle (u, F, F'). In the study of the critical k-sets defined below, two special types of subsets of F play a basic role. These types will be designated by A and B, respectively. Briefly a set X is of type A if the minimizing properties of (u, F, F') are unaltered when the points exterior to a neighborhood of X are disregarded. More precisely, by a *subset of F of type A relative to (u, F, F')* will be meant a closed subset X of F with the property that, given a neighborhood N of X, there is no primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + V$, where $V = F - FN$. A subset X of F of type A relative to (u, F, F') will be said to be of *type B relative to (u, F, F')* if the set XX' is of type A for every set X' of type A. The phrase "relative to (u, F, F')" will be omitted when no confusion arises by so doing. The set $X = F$ is of type B. Moreover if X' and X'' are subsets of F of type B so also is their intersection $X'X''$. For since X' and X'' are of type B, it is seen successively that the sets XX' and $XX'X''$ are of type A whenever X is of type A. Consequently $X'X''$ is of type B, as was to be proved.

By the *critical k-set C determined by a minimizing k-cycle (u, F, F')* will be meant the intersection of all subsets X of F of type B relative to (u, F, F'). It will be seen in Theorem 5.1 below that the set $X = C$ is a subset of F of type B and hence that the set $X = C$ is the smallest subset of F of type B. In general, the critical k-set C is not a member of the admissible class $\mathfrak{F}$ by which it is determined.

In order to understand more clearly the significance of these special subsets and of the definition of critical k-sets, consider the following simple example. Let S be the interval $0 \leqq x \leqq 3\pi$ and let $\mathfrak{F}$ be the admissible class of sets determined by all inequalities of the form $\sin x \leqq c$, where c is a constant. Denote by F and F' the sets in $\mathfrak{F}$ determined by the inequalities $\sin x \leqq 1$ and $\sin x \leqq 0$, respectively. Using the integers modulo 2 as coefficients, let u be the 1-chain defined by the interval $0 \leqq x \leqq 3\pi$ with $x = 0$ and $x = 3\pi$ as its boundary. The 1-cycle (u, F, F') is then minimizing. A closed subset X of F is of type A if, and only if, it contains one of the points $x' = \pi/2$, $x'' = 5\pi/2$ at which $\sin x = 1$. It is of type B if and only if it contains both x' and x''. The set C composed of x' and x'' is the smallest subset of F of type B and is accordingly the critical 1-set determined by (u, F, F').

The following lemma will be useful.

LEMMA 5.1. *Let (u, F, F') be a non-null minimizing k-cycle and let X be a subset of F of type A relative to (u, F, F'). If V denotes the set of vertices of the k-chain v belonging to a k-cycle (v, F, F'') homologous to (u, F, F'), there is no neighborhood N of X such that the set NV is on a primary subset of F. Moreover the set XV is non-null.*

For suppose that there exists a neighborhood N of X and a primary subset G of F such that $NV \subset G$. Then $V \subset G + Y$, where $Y = F - FN$. Since the sum $F' + F'' + G$ of primary subsets of F is also a primary subset of F, we can suppose that this sum is identical with G. Then $G \supset F' + F''$. From the relations, $u \sim v$ on $F \bmod F' + F''$ and $V \subset G + Y$, it follows that $u \sim 0$ on $F \bmod G + Y$, contrary to our assumption that X is of type A. This proves the first part of the lemma. The second follows from the first, since if XV were null, the set NV would be null for a sufficiently small neighborhood N of X.

THEOREM 5. 1. *The critical k-set C determined by a non-null minimizing k-cycle (u, F, F') is a non-null subset of F of type B relative to (u, F, F'). If V denotes the set of vertices of the k-chain v belonging to a minimizing k-cycle (v, F, F'') homologous to (u, F, F'), the set CV is non-null. In fact, the closure C' of the sum of all sets CV of this type is identical with C.*

In order to show that the set CV described in the theorem is non-null, observe that, by virtue of Lemma 5. 1, the set XV is non-null for every subset X of F of type A. The set XV is compact since V is compact. Since the class of subsets of F of type B is closed relative to intersections, the intersection CV of all sets of the form XV, where X is of type B, is non-null, as was to be proved. The set C is closed since it is the intersection of closed sets.

The remaining statements of the theorem will be established by showing that the set $X = C'$ is of type B. For then C and C', being subsets of each other, must be identical. Suppose, therefore, that the set C' is not of type B. Then there exists a subset X of F of type A such that the set XC' is not of type A. We may accordingly choose a neighborhood N of XC' such that $u \sim 0$ on $F \bmod F'' + Y$, where $Y = F - FN$. By virtue of Lemma 2. 2 we can suppose that the set U of vertices of u is on $F'' + Y$. The set NU is then on a primary subset of F. Moreover, $C'U = CU$ in view of our definition of C'. Since U is compact and the set $XC'U = XCU$ is the product of all sets of the form $XX'U$, where X' is a subset of F of type B, there is a subset X' of F of type B such that $XX'U \subset NU \subset F''$. The set $XX'U$ is, therefore, on the primary subset F'' of F, which is impossible, by Lemma 5. 1, since the set XX' is of type A, in view of the fact that X is of type A and X' of type B. Our assumption that the set C' is not of type B is false and the theorem is established.

COROLLARY 1. *If C is the critical k-set determined by a minimizing k-cycle (u, F, F') and $F' \subset E = F - C$, then $u \not\sim 0$ on $F \bmod E$. The relation $F' \subset E$ holds if no primary subset of F has a point in common with C.*

For if $u \sim 0$ on $F \bmod E$ one could suppose, by Lemma 2.2, that u is on $F' + E = E$ and hence would have no vertices in C, contrary to the second statement in Theorem 5.1.

COROLLARY 2. *Homologous minimizing k-cycles determine the same critical k-sets. In fact, if two minimizing k-cycles are homologous to the same minimizing k-cycle, they determine the same critical k-set.*

The critical k-set C belonging to a minimizing k-cycle (u, F, F') is determined completely by the k-cycles on a neighborhood of C, as is shown in the following:

THEOREM 5.2. *Let N be a neighborhood of the critical k-set determined by a non-null minimizing k-cycle (u, F, F'). There exists a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + FN$. In fact, there exists a k-cycle v on $FN \bmod F''N$ such that the system (v, F, F'') defines a minimizing k-cycle homologous to (u, F, F'). If V denotes the set of vertices of a k-chain v of this type, the closure C' of the sum of all sets of the form CV is identical with C.*

In order to prove the first two conclusions in the theorem observe that the set $X = F - FN$, having no points in common with C, cannot be of type A. There exists accordingly a neighborhood N' of X and a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + Y$, where $Y = F - FN'$. By Lemma 2.2 there exists a k-cycle v on $Y \bmod F''Y$ such that $u \sim v$ on $F \bmod F''$. Since $Y \subset FN$ the k-chain v is a k-cycle on $FN \bmod F''N$. Moreover, the system (v, F, F'') is a k-cycle homologous to (u, F, F') and, by Lemma 3.2, it is minimizing. In view of the fact that v is on FN and $u \sim v$ on $F \bmod F''$ it follows that $u \sim 0$ on $F \bmod F'' + FN$.

It remains to prove the last statement in the theorem. As in the proof of Theorem 5.1 it is sufficient to show that the set C' described in Theorem 5.2 is a subset of F of type B. Suppose that the contrary is true. Then there is a subset X of F of type A such that the set CX' is not of type A. Let N' be a neighborhood of XC' and let G be a primary subset of F such that $u \sim 0$ on $F \bmod G + Y'$, where $Y' = F - FN'$. By Lemma 2.2 we can suppose that the set U of vertices of u is on $G + Y'$. The remainder of the proof is like that of Theorem 5.1 provided that we can show that the k-chain $u = (u_n)$ can be chosen so that u is a k-cycle on $FN \bmod F''N$, where F'' is a primary subset of F. To this end let $v = (v_n)$ be a k-chain related to u as described in the last paragraph. Then there is a primary subset F'' of F such that v is a k-cycle on $FN \bmod F''N$ and $u \sim v$ on $F \bmod F''$. We can suppose that

$F'' \supset F' + G$. Let $w = (w_n)$ be a $(k+1)$-chain on F such that the k-chain $w' - u + v$ is on F'' and select a neighborhood N'' of C' such that the closures of N' and N'' lie respectively in N'' and N. From the definition of k-cycles the sequence $u = (u_n)$ can be replaced by any infinite subsequence. Consequently we can suppose that every simplex in the vertex chains u_n, v_n, w_n $(n = 1, 2, \cdots)$ having a vertex in N'' is in N and that every simplex having a vertex in N' is in N''. Let $\bar{u} = (\bar{u}_n)$ be the k-chain obtained from $u = (u_n)$ by deleting the k-simplices in u_n that are exterior to N''. Similarly let $\bar{w} = (\bar{w}_n)$ be the $(k+1)$-chain obtained from $w = (w_n)$ by deleting the $(k+1)$-simplices exterior to N''. The chains $\bar{u}, \bar{w}$ are on FN, are identical with u, w on FN' and are such that the k-chain $\bar{w}' - \bar{u} + v$ is on $F'' + E$, where $E = F(N - N'')$. Let $\bar{v} = (\bar{v}_n)$ be the k-chain obtained from $\bar{w}' - \bar{u} + v$ be deleting all k-simplices of $\bar{w}'_n - \bar{u}_n + v_n$ having no vertex in E. The k-chain $\bar{v}$ has no vertex in N' since, by construction, no simplex of the vertex k-chains u_n, v_n, w_n has vertices in N' and in E. It follows that the vertices of $\bar{u} + \bar{v}$ in N' are vertices of $\bar{u}$ and hence of u. Since u is on $G + Y'$ and $F'' \supset G$ it follows that $\bar{u} + \bar{v}$ is on $F''N + Y'N$. Moreover, the k-chain $\bar{w}' - (\bar{u} + \bar{v}) + v$ is on $F''N$. Hence $\bar{u} + \bar{v} \sim v$ on $FN \bmod F''N$. The k-chain $\bar{u} + \bar{v}$ is accordingly a k-cycle on $FN \bmod F''N$ since v has this property. It is clear that $u \sim \bar{u} + \bar{v}$ on $F \bmod F'$. Replacing u by $\bar{u} + \bar{v}$ we obtain the desired result, namely, that u can be chosen to be on a set $F'' + Y$ in such a way that u is a k-cycle on $FN \bmod F''N$ and F'' is a primary subset of F. This completes the proof of Theorem 5. 2.

THEOREM 5. 3. *No neighborhood N of a point x in the critical k-set C determined by a minimizing k-cycle (u, F, F') is such that the set FN is on a primary subset F'' of F.*

This result follows from Theorems 7. 1 and 7. 2 below. A direct proof can be made as follows. Let x be a point of F for which there exists a neighborhood N of x such that FN is on a primary subset F'' of F. If X is a subset of F of type B, it is easily seen that the closure X' of the set $X - F''X$ is also of type B so that $C \subset X'$. Taking $X = C$ one has $X' = C$. The point x therefore cannot be in C and the theorem is proved.

As an application of the last theorem we have the following:

COROLLARY. *Let $\mathfrak{F}$ be the admissible class of sets $f(x) \leqq c$ determined by a real single valued lower semi-continuous function $f(x)$ on S, as described in* **3**. *Let C be the critical k-set determined by a non-null minimizing k-cycle (u, F, F') and denote by b the least upper bound of $f(x)$ on F. The upper* *limit of* $f(x)$ *on* F *at a point* x' *in* C *is equal to* b. *If* $f(x)$ *is continuous, then* $f(x) = b$ *on* C. *If* $f(x) = b$ *on* C, *then* $F' \subset E = F - C$ *and* $u \not\sim 0$ *on* F *mod* E.

The first conclusion in the corollary follows at once from Theorem 5.3 since if the upper limit of $f(x)$ on F at a point x' in C were not equal to b there would exist a neighborhood N of x' and a constant $c < b$ such that FN would be on the primary subset F'' of F determined by the inequality $f(x) \leqq c$. The last statement in the lemma follows from Corollary 1 to Theorem 5.1.

It is desirable to obtain simple criteria for a point to form a critical k-set. To this end we have the following:

THEOREM 5.4. *A point* x' *forms a critical* k-*set if and only if there exists a* k-*cycle* (u, F, F') *such that given a neighborhood* N *of* x' *then* (1) *there is a primary subset* F'' *of* F *such that* $u \sim 0$ *on* F *mod* $F'' + FN$; (2) *there is no primary subset* F'' *of* F *such that* $u \sim 0$ *on* F *mod* $F' + F'' + Y$, *where* $Y = F - FN$.

The necessity of condition (1) follows from Theorem 5.2 with $C = x'$. The condition (2) is the condition that the set $C = x'$ be of type A and hence also is a necessary condition. To prove the converse observe that when condition (2) holds the k-cycle (u, F, F') must be minimizing, by Lemma 3.1. This condition also implies that any closed subset X of F containing x' is of type A. On the other hand every subset X of F of type A contains x'. If this were not so, one could select a neighborhood N of x' and a neighborhood N' of X having no points in common. By condition (1) it would follow that there exists a primary subset F'' of F such that $u \sim 0$ on F mod $F'' + FN$ and hence such that $u \sim 0$ on F mod $F'' + Y'$, where $Y' = F - FN'$. This would contradict our choice of X as being of type A. Hence $X \supset x'$ and the set $C = x'$ is of type B. It is accordingly the critical k-set determined by (u, F, F') and the theorem is established.

COROLLARY. *In order that a point* x' *form a critical* k-*set it is sufficient that there exist a* k-*cycle* (u, F, F') *such that* (1) *given a neighborhood* N *of* x' *there is a primary subset* F'' *of* F *such that* $u \sim 0$ *on* F *mod* $F'' + FN$; (2) *there exists no primary subset* F'' *of* F *such that* x' *is on* $F' + F''$; (3) $u \not\sim 0$ *on* F *mod* E, *where* E *is the set obtained from* F *by deleting the point* x'. *These conditions are necessary conditions for the point* x' *to form the critical* k-*set determined by a minimizing* k-*cycle* (u, F, F') *if there is no primary subset of* F *containing* x'.

The conditions in the corollary obviously imply those in the last theorem.

The last statement in our corollary follows from Corollary 1 to Theorem 1 since our hypotheses imply that $F' \subset E$.

6. Sum of two critical k-sets. Let C_1 and C_2 be the critical k-sets determined by two minimizing k-cycles (u, F, F') and (v, G, G') with $F = G$. If the k-cycle $(u + v, F, F' + G')$ is also minimizing, one might expect that its critical k-set C would be a subset of $C_1 + C_2$. That this is not always the case can be seen by an example. It is true, however, that $C = C_1 + C_2$ when C_1 and C_2 are at a positive distance apart, as is seen in the following:

LEMMA 6.1. *Suppose that the critical k-sets C_1 and C_2, determined by two non-null minimizing k-cycles (u, F, F') and (v, G, G') with $F = G$, are at a positive distance from each other. Then the k-cycle $(u + v, F, F' + G')$ is minimizing and its critical k-set C is identical with $C_1 + C_2$.*

We can suppose, for convenience, that $F' = G'$. Let N_1 and N_2 be neighborhoods of C_1 and C_2, respectively, whose closures have no point in common. By Theorem 5.2 we can suppose that u is on N_1 and v is on N_2. If the k-cycle $(u + v, F, F')$ were not minimizing, there would exist, by Lemma 3.1, a primary subset F'' of F such that $u + v \sim 0$ on $F \bmod F''$. The k-cycles (u, F, F'') and $(-v, F, F'')$ would be homologous minimizing k-cycles determining disjoint critical k-sets C_1 and C_2, contrary to the result described in Corollary 2 of Theorem 5.1. It follows that the k-cycle $(u + v, F, F')$ is minimizing.

As a second step in the proof of Lemma 6.1 it will be shown that a set X is of type A relative to $(u + v, F, F')$ if, and only if, it is of type A relative to (u, F, F') or (v, F, F'). To do so let N be a neighborhood of X and set $Y = F - FN$. If there is a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + Y$ and $v \sim 0$ on $F \bmod F'' + Y$, then $u + v \sim 0$ on $F \bmod F'' + Y$. Consequently, if X is not of type A relative to either (u, F, F') or (v, F, F') it cannot be of type A relative to $(u + v, F, F')$. Conversely suppose that X is not of type A relative to $(u + v, F, F')$ but is of type A relative to one of the two other k-cycles, say (u, F, F'). Then since C_1 is of type B relative to (u, F, F'), the set XC_1 is of type A relative to (u, F, F') and we can suppose that X has been replaced by XC_1. Since X is not of type A relative to $(u + v, F, F')$ there is a neighborhood N of X and a primary subset F'' of F such that $u + v \sim 0$ on $F \bmod F'' + Y$, where $Y = F - FN$. Since $X \subset C_1$ we can choose N so that N is interior to the neighborhood N_1 described in the last paragraph. But since v is on Y, it follows that $u \sim 0$ on $F \bmod F'' + Y$, contrary to our assumption that X is of type A relative to

(u, F, F'). This proves the statement made at the beginning of this paragraph.

In order to prove that $C \subset C_1 + C_2$ consider a set X of type A relative to $(u + v, F, F')$. Then X must be of type A relative to either (u, F, F') or (v, F, F'), say the former. Since $C_1 + C_2$ is of type B relative to (u, F, F'); the set $X(C_1 + C_2)$ is of type A relative to (u, F, F') and hence relative to $(u + v, F, F')$. Consequently the set $C_1 + C_2$ is of type B relative to $(u + v, F, F')$ and $C \subset C_1 + C_2$. In order to show that $C \supset C_1$ let X be a set of type A relative to (u, F, F'). Then XC_1 is of type A relative to (u, F, F') and hence also relative to $(u + v, F, F')$. The set XC_1C must accordingly be of type A relative to $(u + v, F, F')$ and hence relative to either (u, F, F') or (v, F, F'). Being in N_1 it must be of type A relative to (u, F, F'), that is, the set C is of type B relative to (u, F, F'). Consequently, $C \supset C_1$. By symmetry $C \supset C_2$ so that $C = C_1 + C_2$. This completes the proof of Lemma 6.1.

LEMMA 6.2. *Let C be the critical k-set determined by a non-null minimizing k-cycle (u, F, F') and suppose that C is the sum of two non-null sets C_1 and C_2 at a positive distance from each other. There exist minimizing k-cycles (v, F, F'') and (w, F, F'') whose critical k-sets are C_1 and C_2, respectively, and are such that $(u, F, F') \sim (v + w, F, F'')$.*

In order to prove this result let N_1 and N_2 be neighborhoods of C_1 and C_2 whose closures have no point in common. Let $N = N_1 + N_2$. By Theorem 5.2 we can suppose that u is a k-cycle on $FN \bmod F'N$. Then u is the sum of two k-chains $v + w$ such that v is on N_1 and w is on N_2. Since u is a k-cycle on $FN \bmod F'N$ and N_1, N_2 are disjoint it follows that v is a k-cycle on $FN_1 \bmod F'N_1$. Similarly w is a k-cycle on $FN_2 \bmod F'N_2$. Consequently these systems (v, F, F') and (w, F, F') are k-cycles. Suppose that one of them, say the latter, is not minimizing. Then there exists a primary subset F'' of F such that $w \sim 0$ on $F \bmod F''$. Let X be a subset of F of type A relative to (u, F, F'). The set C_1, being a proper subset of C, is not of type B relative to (u, F, F'). Consequently we can select X so the set C_1X is not of type A. There is, accordingly, a neighborhood $N' \subset N_1$ of C_1X and a primary subset F''' of F such that $u \sim 0$ on $F \bmod F''' + Y'$, where $Y' = F - FN'$. Consequently $v \sim 0$ on $F \bmod F''' + Y$, where $Y = Y' - FN_2$. But this implies that $u = v + w \sim 0$ on $F \bmod F'' + F''' + Y$, contrary to the fact that XC is of type A relative to (u, F, F'). The k-cycles (v, F, F') and (w, F, F') are accordingly minimizing k-cycles. The fact that C_1 and C_2 are the critical k-sets determined by (v, F, F') and (w, F, F') follows readily from Lemma 6.1. This proves Lemma 6.2.

7. Homotopic critical points. In Section 5 critical points were defined by means of homologies. In the present section it will be seen that these critical points are instances of a larger class of critical points that are determined by a special class of homotopic deformations, which we shall call F-deformations. In order to define F-deformations let us first recall the definition of homotopic deformations. Let P be a subset of S. To each point x in P and each value t on $0 \leqq t \leqq 1$ let there correspond a unique point $y(x, t)$ on S. The point function $y(x, t)$ $(0 \leqq t \leqq 1)$ will be said to define a *homotopic deformation* Δ *of* P *on* S if $y(x, 0) = x$ for each point x in P and the function $y(x, t)$ is continuous in its arguments. It will be called an *F-deformation* Δ *of* P if it has the further property that given a point x in P and a value t' on $0 \leqq t \leqq 1$, the points $y(x, t)$ $(t' \leqq t \leqq 1)$ belong to each set F in $\mathfrak{F}$ that contains the point $y(x, t')$. In the sequel it will be convenient to call the set formed by the points $y(x, 1)$, where x is in P, the *image* ΔP *of* P under the deformation Δ. Similarly, for a fixed value of t on $0 \leqq t \leqq 1$, the set of points of the form $y(x, t)$, where x is in P, will be denoted by the symbol $\Delta_t P$.

Consider now a set F in $\mathfrak{F}$. A point x' in F will be called a *homotopic critical point of* F if there does not exist, for any neighborhood N of x', an F-deformation Δ of FN such that for each value t on $0 < t \leqq 1$ the image $\Delta_t(FN)$ under Δ belongs to a primary subset F'' of F. The primary subset of F to be chosen depends on the value of t. This definition is analogous to one given by Morse (XVII, p. 419) and was suggested by it. The homotopic critical points of F are determined completely by F-deformations of the whole set F, as is readily seen by the use of the following:

LEMMA 7.1. *Let x' be a point on a set F in $\mathfrak{F}$ and suppose that x' is not a homotopic critical point of F. There exists a neighborhood N' of x' and an F-deformation Δ of F under which the points exterior to N' are held fast, while for each value t on $0 < t \leqq 1$ the image $\Delta_t P$ under Δ of a compact subset P of FN' is on a primary subset F_t of F.*

In order to prove this result let N be a neighborhood of the point x' for which there exists an F-deformation Δ of FN such that the set $\Delta_t(FN)$ is on a primary subset of F for each fixed value of t on $0 < t \leqq 1$. Such a neighborhood exists since x' is not a homotopic critical point of F. Let N' be an e-neighborhood of x' whose closure is in N and select a real single-valued continuous function $h(x)$ such that $h(x) = 0$ exterior to N' and $0 < h(x) \leqq 1$ on N'. For each value of t on $0 \leqq t \leqq 1$ set $z(x, t) = x$ exterior to N' and let $z(x, t) = y[x, h(x)t]$ on FN', where $y(x, t)$ is the point function defining the F-deformation Δ. It is clear that the function $z(x, t)$ $(0 \leqq t \leqq 1)$ defines an F-deformation Δ' of F in which the points exterior to N' are held fast. Consider now a compact subset P of FN' and let m be the minimum of $h(x)$ on P. Since $h(x) > 0$ on P we have $m > 0$. Given a value t on $0 < t \leqq 1$ let F' be a primary subset of F such that $\Delta_{mt}(FN)$ is on F'. For each point x in P, the point $y(x, mt)$ must be on F' since $P \subset FN$. From the definition of F-deformations and the relation $h(x) \geqq m$ on P it follows that for each point x in P the point $z(x, t) = y[x, h(x)t]$ is also on F'. Consequently, the set $\Delta'_t P$ is on F' and the lemma is established.

COROLLARY. *The neighborhood N' of x' and the F-deformation of F described in Lemma 7.1 can be chosen so that for each value of t on $0 \leqq t \leqq 1$ the set $\Delta_t(FN')$ is on a prescribed neighborhood N of x'.*

It is clear from the proof of Lemma 7.1 that the neighborhood N' can be chosen so that its closure is in the prescribed neighborhood of x'. Moreover, the function $h(x)$ there used can be replaced by $rh(x)$, where r is a positive constant less than unity. When this has been done, it follows from continuity considerations that, for sufficiently small values of r, the F-deformation Δ will have the property described in the corollary.

THEOREM 7.1. *If C is the critical k-set determined by a non-null minimizing k-cycle (u, F, F'), then each point of C is a homotopic critical point of the set F belonging to (u, F, F').*

In order to prove this result suppose that there exists a point x' in C that is not a homotopic critical point relative to F. Let N' be a neighborhood of x' and let Δ be an F-deformation of F related to x' as described in Lemma 7.1. Let N'' be a neighborhood of x' whose closure is in N' and set $X = C - CN''$. In view of the definition of C, the set X cannot be of type B relative to (u, F, F'). Consequently there exists a subset X' of F of type A such that the set XX' is not of type A. There exists, accordingly, a neighborhood N^* of XX' and a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + Y^*$, where $Y^* = F - FN^*$. By Lemma 2.2 we can suppose that the set U of vertices of u is on $F'' + Y^*$. Let N be a neighborhood of CX' whose closure is interior to the neighborhood $N^* + N'$ of CX'. We shall show that there exists a primary subset G of F such that the image ΔU of U under the F-deformation Δ is on $G + Y$, where $Y = F - FN$. To this end let $U'' = F''U$ and let U' be the closure of $U - U''$. Since Δ is an F deformation $\Delta U'' \subset F''$. The points of $Q = U' - U'N'$ are held fast under Δ and are at a positive distance from FN since Q is exterior to $N^* + N'$. There is, accordingly, a neighborhood N_1 of Q such that $\Delta(U'N_1) \subset Y$. Finally the compact set $P = U' - U'N_1$ is on N' and, by Lemma 7.1, there exists a primary subset G of F such that $\Delta P \subset G$. We can suppose $G \supset F''$. Then $\Delta U = \Delta U'' + \Delta(U'N_1) + \Delta P$ is on $G + Y$, as was to be proved. Since the boundary u' is deformed on F' under Δ it follows from the theory of homotopic deformations that $u \sim 0$ on F mod $F' + G + Y$, contrary to the fact that C is a subset of F of type A, by Theorem 5.1. Our assumption that there exists a point x' in C that is not a homotopic critical point relative to F is accordingly false and Theorem 7.1 is established.

THEOREM 7.2. *There is no neighborhood N of a homotopic critical point x' of a set F in $\mathfrak{F}$ such that the set FN is on a subset of F. In particular a point x' in the critical k-set determined by a minimizing k-cycle (u, F, F') has this property.*

For suppose there is a neighborhood N of a homotopic critical point of F such that FN is on a proper subset F'' of F in $\mathfrak{F}$. Let Δ be the F-deformation of F in which all points of F are held fast. Then for each value t on $0 \leqq t \leqq 1$ the image $\Delta_t(FN)$ of FN is identical with FN and hence is on F''. This contradicts the definition of homotopic critical points of F. The first statement is accordingly established. The second follows from the fact that each point of the critical k-set C of a minimizing k-cycle (u, F, F') is a homotopic critical point of F.

Recalling the definition of admissible classes $\mathfrak{F}$ determined by a function $f(x)$, one obtains from the last theorem the following further result.

COROLLARY 1. *Let F be a set in the admissible class $\mathfrak{F}$ determined by a real single valued function $f(x)$, as described in* **3**. *Let b be the least upper bound of $f(x)$ on F. At each homotopic critical point x', the upper limit of $f(x)$ on F is equal to b. If $f(x)$ is continuous at x', then $f(x') = b$.*

COROLLARY 2. *Let $\mathfrak{F}$ be the admissible class determined by a real single valued lower semi-continuous function as described in* **3**. *Let C be the critical k-set determined by a minimizing k-set (u, F, F') and let b be the least upper bound of $f(x)$ on F. If there exists an F-deformation Δ of F such that $f(x)$ is continuous on ΔF, then $f(x) = b$ on C.*

8. Two lemmas. In the present section further properties of F-deformations will be developed. The results obtained will be useful in the next section. Let $P \subset Q$ be subsets of S and let Δ be an F-deformation of P defined by functions $y(x, t)$ $(0 \leqq t \leqq 1)$. The set P will be said to be *deformed on Q* under Δ

if the point $y(x, t)$ is on Q for each point x in P and each value t on $0 \leqq t \leqq 1$. The following lemma will be useful.

LEMMA 8.1. *Let F be a set in $\mathfrak{F}$ and let P be a compact subset of F containing no homotopic critical points of F. Given a neighborhood N of P there exists a neighborhood N^* of P and an F-deformation Δ of F having the following properties: The points of F exterior to N^* are held fast under Δ. The points of FN^* are deformed on FN under Δ. Finally, the image ΔP of P under Δ is on a primary subset F'' of F.*

In order to prove this result let N be a neighborhood of P. To each point x' of P let there correspond a neighborhood N' of x' and an F-deformation Δ related to x' and N' as described in Lemma 7.1 and its corollary. Since P is compact we can select a finite number of these neighborhoods $N_1, \cdots, N_q$ covering P. Let $\Delta_1, \cdots, \Delta_q$ be the corresponding F-deformations of F. Let Δ be the F-deformation of F obtained by applying successively the F-deformations $\Delta_1, \cdots, \Delta_q$. The F-deformation Δ and the neighborhood $N^* = N_1 + \cdots + N_q$ of P have the properties described in the lemma. The points exterior to N^* are held fast under each of the deformations $\Delta_1, \cdots, \Delta_q$ and hence also under Δ. Moreover, the set FN^* is deformed on FN under Δ since each of the deformations $\Delta_1, \cdots, \Delta_q$ has this property. It remains to show that ΔP is on a primary subset F'' of F. To do so we recall that under the F-deformation Δ, the points of $P - PN_1$ are held fast. There exists, accordingly, a neighborhood N_1'' of $P - PN_1$ such that the closure Q_2 of the image $\Delta_1(PN_1'')$ is on $N_2' = N_2 + \cdots + N_q$. By Lemma 7.1 with P replaced by $\dot{P} = PN_1''$ there is a primary subset F_1 of F such that the set $Q_1 = \Delta_1(P - PN_1'')$ is on F_1. The set $P_1 = \Delta_1 P$ is accordingly the sum of two compact sets Q_1 and Q_2 such that Q_1 is on a primary subset F_1 on F and Q_2 is on N_2'. Repeating the argument made in the last four sentences with P replaced by Q_2, Δ_1 replaced by Δ_2, and N'_2 replaced by $N'_3 = N_3 + \cdots + N_q$, it is found that the set $\Delta_2 Q_2$ is the sum of two compact sets Q'_1 and Q'_2 such that Q'_1 is on a primary subset F'_1 of F and Q'_2 is on N'_3. The set $\Delta_2 Q_1$ is on F_1 since Δ_2 is an F-deformation. In the next repetition replace P by Q'_2, Δ_2 by Δ_3 and N'_2 by $N'_4 = N_4 + \cdots + N_q$. Proceeding in this manner it is found that the image ΔP of P under Δ is on a primary subset $F'' = F_1 + F'_1 + \cdots$ of F. This completes the proof of Lemma 7.2.

The proof of Theorem 9.2 is based on the following:

LEMMA 8.2. *Let F be a non-null set in $\mathfrak{F}$ and let H be an isolated set of homotopic critical points relative to F. Let N be a neighborhood of H*

containing in its closure no homotopic critical point of F that is not in H. Consider a k-cycle u on FN mod $F'N$, where F' is a primary subset of F. If $u \sim 0$ on F mod F', there is a primary subset F'' of F such that $u \sim 0$ on FN mod $(F' + F'')N$. The k-cycle (u, F, F') is minimizing if and only if there is no primary subset F'' of F such that $u \sim 0$ on FN mod $(F' + F'')N$.

In order to prove the first conclusion in the lemma let N', N'' be neighborhoods of H such that the closures of N', N'' are in N'', N respectively. Set $E = F(N'' - N')$. Given a k-cycle $u = (u_n)$ on FN mod $F'N$ such that $u \sim 0$ on F mod F', we shall show first that there is a k-cycle v on FN mod FN such that $u \sim v$ on FN mod $F'N$ and v is on $F'N + E$. To this end let $w = (w_n)$ be a $(k+1)$-chain such that $w' - u$ is on F'. We may replace the sequences (u_n), (w_n) by subsequences, if necessary, so that every simplex of u_n or of w_n having a point in N' is on N''. Let $\bar{w} = (\bar{w}_n)$ be the $(k+1)$-chain obtained from w by deleting all $(k+1)$-simplices having no point in common with N'. Then the k-chain $v = \bar{w}' - u$ is on $F'N + E$ and $u \sim v$ on FN mod $F'N$. Moreover, v is a k-cycle on FN mod $F'N$ since u has this property. Let V be the vertices of v and let P be the closure of the set $V - VF'$. If P is null, then $u \sim 0$ on FN mod $F'N$ and the first conclusion in the lemma holds with $F'' = 0$. If $P \neq 0$, then P is compact and contains no homotopic critical points since $P \subset N - N'$. Let Δ be an F-deformation related to P and N as described in the last lemma. Under Δ the points of F in a neighborhood N' of P are deformed on FN while the remaining points are held fast. Hence the set V is deformed on FN under Δ. Moreover, by Lemma 8.1, there is a primary subset F'' of F such that $\Delta P \subset F''$. The set $\Delta(V - P)$ is on F' since $V - P$ is on F'. Hence the set V is deformed under Δ on FN onto the set $F'N + F''N$. Consequently, $v \sim 0$ on FN mod $F'N + F''N$ and since $u \sim v$ on FN mod $F'N$ we have also $u \sim 0$ on FN mod $F'N + F''N$, as was to be proved.

The last statement in the lemma follows from the one just established. If the k-cycle (u, F, F') is not minimizing, there exists, by Lemma 3.1, a primary subset F'' of F such that $u \sim 0$ on F mod $F' + F''$. We have just seen that in this case F'' can be chosen so that $u \sim 0$ on FN mod $(F' + F'')N$. Conversely if there is a primary subset of F such that $u \sim 0$ on FN mod $(F' + F'')N$ then $u \sim 0$ on F mod $F' + F''$ and (u, F, F') cannot be minimizing. This completes the proof of Lemma 8.2.

9. *F*-homology classes and groups. Two minimizing k-cycles will be said to be *F-homologous* if they are homologous to a third minimizing k-cycle.

In view of Lemma 3.3 a minimizing k-cycle (u, F, F') with $F \neq 0$ is F-homologous to a k-cycle (v, G, G') if, and only if, the k-cycle $(u - v, F + G, F' + G')$ is not minimizing, that is, if, and only if, $F = G$ and there is a primary subset F'' of F such that $u \sim v$ on $F \bmod F''$. This last criterion suggests the term "F-homologous."

LEMMA 9.1. *The F-homology relation of minimizing k-cycles is reflexive, symmetric, and transitive.*

It is clear that the F-homology relationship is reflexive and symmetric. In order to prove that it is transitive let (u, F, F'), (v, G, G') and (v, G, G'), (w, H, H') be pairs of F-homologous minimizing k-cycles. Then $F = G = H$. If $F = 0$, then $(u, F, F') = (w, H, H') = (0, 0, 0)$. Suppose, therefore, that $F \neq 0$ and select a proper subset F'' of F in $\mathfrak{F}$ such that $u \sim v$ on $F \bmod F''$ and $v \sim w$ on $F \bmod F''$. Then $u \sim w$ on $F \bmod F''$ and (u, F, F') is F-homologous to (w, F, F'). This proves the lemma.

A class M_k composed of a minimizing k-cycle (u, F, F') and all minimizing k-cycles that are F-homologous to it will be called an *F-homology class.* The F-homology class composed of the minimizing k-cycle $(0, 0, 0)$ will be called the *zero F-homology class* $M_k = 0$. Let M_k be an F-homology class and let (u, F, F') be an k-cycle in M_k. The class containing $(-u, F, F')$ will be denoted by $-M_k$ and we set $M_k + (-M_k) = 0$. If $M'_k \neq -M_k$ is a second F-homology class and (v, G, G') is in M'_k, then the k-cycle $(u + v, F + G, F' + G')$ is a minimizing k-cycle belonging to a class which we define as the sum $M_k + M'_k$ of M_k and M'_k. It is easily seen that this definition of the sum $M_k + M'_k$ is independent of our choice of the k-cycles (u, F, F') and (v, G, G'). We have accordingly the following result.

THEOREM 9.1. *Under the above definition of addition the F-homology classes M_k define an abelian group $\mathfrak{M}_k$, called the k-th F-homology group of $\mathfrak{F}$.*

It is interesting to observe in passing that if F is composed only of the set $F = S$ and $F = 0$, the k-th F-homology group $\mathfrak{M}_k$ of $\mathfrak{F}$ is identical with the k-th homology group $\mathfrak{H}_k$ of S. An admissible class of this type is determined by the function $f(x) =$ constant on S. A relation between $\mathfrak{M}_k$ and $\mathfrak{H}_k$ when $\mathfrak{F}$ is arbitrary will be found in Theorem 10.2 below.

The k-th F-homology group $\mathfrak{M}_k$ is completely determined by the k-cycles on a neighborhood of the homotopic critical points, as is seen in the following:

THEOREM 9.2. *Let H be the set of all homotopic critical points on S determined by $\mathfrak{F}$. Let N be a neighborhood of H and S' the closure of N in S. The class $\mathfrak{F}'$ of all sets of the form FS', where F is in $\mathfrak{F}$, forms an admissible* *class relative to* S'. *Moreover, the* k*-th* F*-homology group* $\mathfrak{M}'_k$ *of* $\mathfrak{F}'$ *is isomorphic to the* k*-th* F*-homology group* $\mathfrak{M}_k$ *of* $\mathfrak{F}$.

To prove this we note that every critical k-set C determined by a minimizing k-cycle (u, F, F') belongs to H, by virtue of Theorem 7.1. Hence N is also a neighborhood of C. By Theorem 5.2, with N replaced by S', there is a k-cycle v on $FS' \bmod F'S'$, such that $u \sim v$ on $F \bmod F'$. The system $(v, FS', F'S')$ is accordingly a minimizing k-cycle determined by the class $\mathfrak{F}'$. Let $(w, FS', F''S')$ be a second minimizing k-cycle determined by F' such that $u \sim w$ on $F \bmod F' + F''$. Then $v \sim w$ on $F \bmod F' + F''$ and by Lemma 8.2 with $u = v - w$, there is a primary subset G of F such that $v \sim w$ on $FN \bmod GN$. The k-cycles $(v, FS', F'S')$ and $(w, FS', F''S')$ belong, therefore, to the same class M'_k in $\mathfrak{M}'_k$. Let the class M'_k correspond to the class M_k in $\mathfrak{M}_k$ containing (u, F, F'). Thus to each element M_k in $\mathfrak{M}_k$ there corresponds a unique element M'_k in $\mathfrak{M}'_k$. Consider next a k-cycle $(u, FS', F'S')$ in a class M'_k of $\mathfrak{M}'_k$. Then by Lemmas 8.2 and 3.1 the k-cycle (u, F, F') is also minimizing. It follows that each class M'_k in $\mathfrak{M}'_k$ is the image of one in $\mathfrak{M}_k$. The correspondence is obviously an isomorphism.

By virtue of Corollary 2 of Theorem 5.1 each homology class M_k determines a unique critical k-set C, namely, the critical k-set determined by one of its members. As a further result we have

THEOREM 9.3. *Let* H *be the set of all homotopic critical points on* S *and* H_1 *an isolated subset of* H. *Set* $H_2 = H - H_1$. *The* F*-homology classes whose critical* k*-sets are in* H_1 *form a subgroup* $\mathfrak{M}_{k1}$ *of* $\mathfrak{M}_k$. *Similarly the* F*-homology classes whose critical* k*-sets are in* H_2 *form a subgroup* $\mathfrak{M}_{k2}$ *and* $\mathfrak{M}_k$ *is the direct sum of* $\mathfrak{M}_{k1}$ *and* $\mathfrak{M}_{k2}$. *The groups* $\mathfrak{M}_{k1}$ *and* $\mathfrak{M}_{k2}$ *are completely determined by the minimizing* k*-cycles over neighborhoods* N_1 *and* N_2 *of* H_1 *and* H_2 *respectively.*

This result follows readily from the last theorem by taking N so small that S' is the sum of two sets S_1, S_2 containing H_1 and H_2 respectively.

10. Analogues of the critical point relations of Morse. The analogue of the critical point relations of Morse given in this section consists of relations between groups $\mathfrak{H}_k$, $\mathfrak{M}_k$, $\mathfrak{N}_k$ and $\mathfrak{P}_k$ $(k = 0, 1, \cdots)$, where $\mathfrak{H}_k$ is the k-th homology group of S, $\mathfrak{M}_k$ is the k-th F-homology group of $\mathfrak{F}$, and $\mathfrak{N}_k$, $\mathfrak{P}_k$ are the subgroups of $\mathfrak{M}_k$ described in Lemma 10.1 below. More specifically, relations between the group $\mathfrak{H}_k$ and the difference groups $\mathfrak{N}_k - \mathfrak{P}_k$ and between the groups $\mathfrak{P}_k$ and $\mathfrak{M}_{k+1} - \mathfrak{N}_{k+1}$ are obtained. It will be seen in the next section that in case the group Λ used in defining chains is a field and $\mathfrak{M}_k$ is of finite dimension, these relations imply that the groups $\mathfrak{H}_k$ and $\mathfrak{N}_k - \mathfrak{P}_k$ are of the same dimension. Similarly, the groups $\mathfrak{P}_k$ and $\mathfrak{M}_{k+1} - \mathfrak{N}_{k+1}$ are of the same dimension. The equality of the dimensions of these groups constitute the critical point relations of Morse, as will be seen in Theorem 11. 1 below.

LEMMA 10. 1. *The set of all F-homology classes M_k in $\mathfrak{M}_k$ containing a k-cycle of the form $(u, F, 0)$ forms a subgroup $\mathfrak{N}_k$ of $\mathfrak{M}_k$. Moreover $\mathfrak{M}_0 = \mathfrak{N}_0$. The set of all F-homology classes M_k in $\mathfrak{N}_k$ containing a k-cycle (u, F, F') such that $u \sim 0$ on S mod F'' for a suitably chosen proper subset F'' of F in $\mathfrak{F}$ forms, with $M_k = 0$, a subgroup $\mathfrak{P}_k$ of $\mathfrak{N}_k$.*

This result follows at once from the definition of the classes $\mathfrak{N}_k$ and $\mathfrak{P}_k$. It should be observed that every k-cycle (u, F, F') belonging to non-null class M_k in $\mathfrak{P}_k$ has the property that $u \sim 0$ on S mod F'' for a suitably chosen proper subset F'' of F in $\mathfrak{F}$. In fact every class M_k in $\mathfrak{M}_k$ containing a k-cycle with this property is necessarily in $\mathfrak{P}_k$.

The set F belonging to a minimizing k-cycle (u, F, F') will be called a minimizing k-set. Two k-cycles in the same F-homology class M_k determine the same minimizing k-set, which we shall denote by $F(M_k)$. We have the following:

THEOREM 10. 1. *Let F be a set in $\mathfrak{F}$. The non-zero F-homology classes in $\mathfrak{M}_k$, $\mathfrak{N}_k$, $\mathfrak{P}_k$ having $F(M_k) = F$ form, with $M_k = 0$, groups $\mathfrak{M}_k(F)$, $\mathfrak{N}_k(F)$, $\mathfrak{P}_k(F)$ respectively. The groups $\mathfrak{M}_k$, $\mathfrak{N}_k$, $\mathfrak{P}_k$, $\mathfrak{M}_k - \mathfrak{N}_k$, $\mathfrak{N}_k - \mathfrak{P}_k$ are direct sums over $\mathfrak{F}$ of the corresponding groups $\mathfrak{M}_k(F)$, $\mathfrak{N}_k(F)$, $\mathfrak{P}_k(F)$, $\mathfrak{M}_k(F) - \mathfrak{N}_k(F)$, $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$. In fact the direct sums need to be taken only over the class $\mathfrak{F}_k$ of all minimizing k-sets F.*

This result is an immediate consequence of the definition of the sum of F-homology classes given in the last section.

In view of the fact that the difference group $\mathfrak{N}_k - \mathfrak{P}_k$ is the direct sum over $\mathfrak{F}$ (or over $\mathfrak{F}_k$) of the difference groups $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$, one can obtain a satisfactory relationship between $\mathfrak{N}_k - \mathfrak{P}_k$ and $\mathfrak{H}_k$ by obtaining a relation between $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ and $\mathfrak{H}_k$ for each set F in $\mathfrak{F}$ (or in $\mathfrak{F}_k$). This will be done in the following:

THEOREM 10. 2. *Let F be a set in $\mathfrak{F}$. Denote by $\mathfrak{H}_k(F)$ the set of all homology classes $\mathfrak{H}_k$ in the k-th homology group $\mathfrak{H}_k$ of S that contain a k-cycle u on F. Similarly denote by $\mathfrak{H}^*_k(F)$ the set of all homology classes in $\mathfrak{H}_k$ that contain a k-cycle on a proper subset of F in $\mathfrak{F}$. The sets $\mathfrak{H}_k(F)$ and $\mathfrak{H}^*_k(F)$ are subgroups of $\mathfrak{H}_k$ and the difference group $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ is isomorphic with the difference group $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ determined by the* *groups* $\mathfrak{N}_k(F)$, $\mathfrak{P}_k(F)$ *described in Theorem* 10.1. *The difference group* $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ *contains only the zero element of* $\mathfrak{H}_k$ *when* F *is not a minimizing* k*-set.*

It is clear that the sets $\mathfrak{H}_k(F)$, $\mathfrak{H}^*_k(F)$ form subgroups of $\mathfrak{H}_k$. In order to prove the remaining result consider first two homology classes H_k and H'_k belonging to the same non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. Let u, v be k-cycles on F belonging to $\mathfrak{H}_k$, $\mathfrak{H}'_k$, respectively. The k-cycles $(u, F, 0)$, $(v, F, 0)$ are minimizing. Suppose, for example, that the k-cycle $(u, F, 0)$ is not minimizing. Then it is homologous to a k-cycle $(w, F', 0)$ such that F' is a proper subset of F in $\mathfrak{F}$. The k-cycle w is in H_k and hence H_k belongs to $\mathfrak{H}^*_k(F)$, contrary to our choice of H_k as belonging to a non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. It follows that the k-cycle $(u, F, 0)$, and hence also $(v, F, 0)$, is minimizing. In view of the fact that the class $H_k - H'_k$ is in $\mathfrak{H}^*_k(F)$, there exists a proper subset F' of F in $\mathfrak{F}$ such that $u - v \sim 0$ on S mod F'. From this result we conclude that the F-homology classes M_k, M'_k containing the k-cycles $(u, F, 0)$, $(v, F, 0)$, respectively, belong to $\mathfrak{N}_k(F)$ and are such that $M_k - M'_k$ is in $\mathfrak{P}_k(F)$. The classes M_k, M'_k accordingly belong to the same coset of $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$. Moreover, this coset is not the zero coset $\mathfrak{P}_k(F)$. Otherwise there would exist a proper subset F' of F in $\mathfrak{F}$ such that $u \sim 0$ on $S \bmod F'$ and the class H_k would be in $\mathfrak{H}^*_k(F)$, contrary to our assumption that H_k belongs to a non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. Since the classes H_k, H'_k need not be distinct, it has been shown that to each non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ there corresponds a unique non-zero coset $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$, related in the manner just described.

In order to show that this correspondence is one-to-one consider two F-homology classes M_k, M'_k belonging to the same non-zero coset of $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ and let $(u, F, 0)$, $(v, F, 0)$ be k-cycles in M_k, M'_k, respectively. Since M_k is not in $\mathfrak{P}_k(F)$ the relation $u \not\sim 0$ on $S \bmod F'$ holds for every proper subset F' of F in $\mathfrak{F}$. Consequently the k-cycle u belongs to an homology class H_k in $\mathfrak{H}_k(F)$ which is not in $\mathfrak{H}^*_k(F)$. Similarly, the k-cycle v belongs to a second homology class H'_k of this type. From the fact that $M_k - M'_k$ is in $\mathfrak{P}_k(F)$, the relation $u - v \sim 0$ on $S \bmod F'$ holds for a suitably chosen proper subset F' of F in $\mathfrak{F}$. It follows that $H_k - H'_k$ is in $\mathfrak{H}^*_k(F)$ and hence that H_k, H'_k belong to the same non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. The correspondence between the non-zero cosets of the difference groups $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ and $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ is therefore one-to-one. If the zero cosets are made to correspond, an isomorphism is obtained, as one readily verifies. This proves Theorem 10.2.

It should be observed that the group $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ is essentially the group of homology classes determined by k-cycles u on F which are non-bounding modulo every proper subset of F in $\mathfrak{F}$.

COROLLARY 1. *Suppose that the sets* $F \neq S$ *in* $\mathfrak{F}$ *are compact. Let* $F \neq G$ *be two subsets of* $\mathfrak{F}$ *such that* $F \supset G$ *and every minimizing* k*-set that is a proper subset of* F *is a subset of* G. *Then* $\mathfrak{H}^*_k(F) = \mathfrak{H}_k(G)$ *and* $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F) = \mathfrak{H}_k(F) - \mathfrak{H}_k(G)$. *Moreover* $\mathfrak{H}_k = \mathfrak{H}_k(S)$.

THEOREM 10. 3. *Suppose that the sets* $F \neq S$ *in* $\mathfrak{F}$ *are compact and let* G *be a set in* $\mathfrak{F}$. *Let* $\mathfrak{P}'_k(G)$ *be the set of all* F*-homology classes* M_k *containing a* k*-cycle* (u, F, F') *such that* $F \subset G$ *and* $u \sim 0$ *on* $G \bmod F''$ *for a suitably chosen proper subset* F'' *of* F *in* $\mathfrak{F}$. *Let* $\mathfrak{P}_k''(G)$ *be the set of all* F*-homology classes* M_k *containing a* k*-cycle* (u, F, F') *for which there exist proper subsets* F'', G'' *of* F, G *in* $\mathfrak{F}$ *such that* $u \sim 0$ *on* $G'' \bmod F''$. *The sets* $\mathfrak{P}_k'(G)$, $\mathfrak{P}_k''(G)$ *are subgroups of* $\mathfrak{P}_k$, $\mathfrak{P}_k'(G)$ *respectively. Moreover, the difference group* $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$ *is isomorphic with the difference group* $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$ *determined by groups* $\mathfrak{M}_{k+1}(G)$, $\mathfrak{N}_{k+1}(G)$ *described in Theorem* 10. 1. *These difference groups consist only of the zero coset when* G *is not a minimizing* $(k+1)$*-set.*

The first two statements in the theorem are immediate. In order to prove the last statement in the theorem let M_k, M_k' be F-homology classes belonging to the same non-zero coset of $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$. Let (u, F, F') and (v, H, H') be k-cycles in M_k, M_k', respectively, such that $u \sim 0$ on $G \bmod F''$, and $v \sim 0$ on $G \bmod H''$, where F'', H'' are proper subsets F, H in $\mathfrak{F}$. By part 1 of Lemma 2. 3 there exist $(k+1)$-cycles w_1, w_2 on $G \bmod F$ such that $u \sim w_1'$ on $F \bmod F''$, and $v \sim w_2'$ on $H \bmod H''$. We can suppose that $u = w_1'$, $v = w_2'$. The $(k+1)$-cycle (w_1, G, F) is minimizing, since otherwise there would exist a proper subset G'' of G in $\mathfrak{F}$ such that $w_1 \sim 0$ on $G \bmod G''$. This would imply that $u \sim 0$ on G'' and hence that M_k is in $\mathfrak{P}_k''(G)$, contrary to our choice of M_k. The F-homology class M_{k+1} containing (w_1, G, F) is in $\mathfrak{M}_{k+1}(G)$ but not in $\mathfrak{N}_{k+1}(G)$. If it were in $\mathfrak{N}_{k+1}(G)$ one could select a proper subset G'' of G and F'' of F in $\mathfrak{F}$, such that $w'_1 = u \sim 0$ on $G'' \bmod F''$, contrary to our choice of u. Similarly the $(k+1)$-cycle (w_2, G, H) is a minimizing $(k+1)$-cycle belonging to an F-homology class M'_{k+1} in $\mathfrak{M}_{k+1}(G)$ which is not in $\mathfrak{N}_{k+1}(G)$. Since the class $M_k - M_k'$ is in $\mathfrak{P}_k''(G)$ there exist proper subsets F''' of $F + H$ and G'' of G in $\mathfrak{F}$ such that $u - v \sim 0$ on $G'' \bmod F'''$. From this property it follows readily that the classes M_{k+1}, M'_{k+1} belong to the same non-zero coset in $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$. Thus, to each non-zero coset of $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$ there corresponds a unique non-zero coset of $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$. In order to show that this correspondence is one to one, consider F-homology classes M_{k+1}, M'_{k+1} belonging to a non-zero coset of $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$ and let (w_1, G, G'), (w_2, G, G'') be minimizing $(k+1)$-cycles in M_{k+1}, M'_{k+1} respectively. By Theorem 4.2, with $G = S$, one can assume that the k-cycles $(w_1', G', 0)$, $(w_2', G'', 0)$ are minimizing. These k-cycles then belong to F-homology classes M_k, M_k' in $\mathfrak{P}_k'(G)$. If M_k were in $\mathfrak{P}_k''(G)$, there would exist a proper subset G''' of G in $\mathfrak{F}$ such that $w_1' \sim 0$ on G''' mod G' and M_{k+1} would be in $\mathfrak{N}_{k+1}(G)$, which is not the case. Similarly, M_k' does not belong to $\mathfrak{P}_k''(G)$. On the other hand $M_k - M_k'$ is in $\mathfrak{P}_k''(G)$ since otherwise $M_{k+1} - M'_{k+1}$ could not be in $\mathfrak{N}_{k+1}(G)$, as one readily verifies. The correspondence between the groups $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$ and $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$ is accordingly one to one, when zero cosets are made to correspond. It is easily seen that the correspondence is an isomorphism. This completes the proof of Theorem 10.3.

COROLLARY. *Let $G \supset G'$ be sets in $\mathfrak{F}$ and suppose that every minimizing $(k+1)$-set contained in G is in G'. Then $\mathfrak{P}_k''(G) = \mathfrak{P}_k'(G')$ and the group $\mathfrak{P}_k'(G) - \mathfrak{P}_k'(G')$ is isomorphic with the group $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$, provided the sets $F \neq S$ in $\mathfrak{F}$ are compact. Moreover, $\mathfrak{P}_k'(S) = \mathfrak{P}_k$.*

11. The case when Λ is a field. In this section it is assumed that the group Λ is a field. The critical point relations of Morse can then be obtained in their original form. This is done in Theorem 11.1 below. First we make the following definitions and observations. If $u = \Sigma\lambda_i z_i$ is a vertex k-chain, we set $\lambda u = \Sigma(\lambda\lambda_i)z_i$. Given a Vietoris k-chain $u = (u_n)$ we define $\lambda u = (\lambda u_n)$. If (u, F, F') is a minimizing k-cycle, so also is $(\lambda u, F, F')$ when $\lambda \neq 0$, and they determine the same critical k-sets. Let M_k be an F-homology class containing (u, F, F'). By λM_k $(\lambda \neq 0)$ will be meant the class containing $(\lambda u, F, F')$. If $\lambda = 0$ we set $\lambda M_k = 0$. The relation $\lambda(M_k + M'_k) = \lambda M_k + \lambda M'_k$ holds for every pair of classes M_k and M'_k in $\mathfrak{M}_k$. The groups described in the last two sections, as well as the k-th homology group $\mathfrak{H}_k$, are distributive with respect to λ in this manner. The number of linearly independent elements (coefficients in Λ) in $\mathfrak{M}_k$ in a maximal set will be called the dimension m_k of $\mathfrak{M}_k$. The dimensions of the groups $\mathfrak{P}_k$ and $\mathfrak{N}_k - \mathfrak{P}_k$ will be denoted by p_k, r_k respectively. Suppose that m_k is finite when $k \leqq q$. Then for a fixed value of $k \leqq q$ for which $m_k > 0$ there exists a finite number of non-zero minimizing k-sets $F_1 \subset F_2 \cdots \subset F_n$. The groups $\mathfrak{M}_k(F_i)$, $\mathfrak{P}_k(F_i)$, $\mathfrak{N}_k(F_i) - \mathfrak{P}_k(F_i)$ have dimensions m_{ki}, p_{ki}, r_{ki}, respectively, and

$$m_k = m_{k1} + \cdots + m_{kn}, \quad p_k = p_{k1} + \cdots + p_{kn}, \quad r_k = r_{k1} + \cdots + r_{kn}.$$

By Theorem 10.2 and its corollary the dimension of the group $\mathfrak{H}_k(F_i) - \mathfrak{H}_k(F_{i-1})$ $(i > 0)$ is equal to r_{ki} and the dimension of $\mathfrak{H}_k(F_1)$ is r_{k1} provided the sets $F \neq S$ in $\mathfrak{F}$ are compact. Consequently, the dimension of the k-th homology group is equal to $r_{k1} + r_{k2} + \cdots + r_{kn'} = r_k$. Similarly, by Theorem 10.3 and its corollary the dimension of the difference group $\mathfrak{P}'_{k-1}(F_i) - \mathfrak{P}'_{k-1}(F_{i-1})$ $(k > 0)$ is equal to the dimension $s_{ki} = m_{ki} - p_{ki} - r_{ki}$ of $\mathfrak{M}_k(F_i) - \mathfrak{R}_k(F_i)$ if $i > 1$ and the sets $F \neq S$ in $\mathfrak{F}$ are compact, as we shall assume. Moreover, the dimension of $\mathfrak{P}'_{k-1}(F_1)$ is equal to $s_{k1} = m_{k1} - p_{k1} - r_{k1}$. It follows that the dimension p_{k-1} of $\mathfrak{P}_{k-1}$ is equal to the dimension $s_{k1} + \cdots + s_{kn} = m_k - p_k - r_k$ of $\mathfrak{M}_k - \mathfrak{R}_k$. Hence $m_k = r_k + p_k + p_{k-1}$ $(0 < k \leqq q)$. Since $\mathfrak{M}_0 \equiv \mathfrak{R}_0$ we have $m_0 = r_0 + p_0$. This proves the following.

THEOREM 11.1. *Suppose that the sets $F \neq S$ in $\mathfrak{F}$ are compact. Let q be a positive integer and suppose that when $k \leqq q$ the dimension m_k of the k-th F-homology group $\mathfrak{M}_k$ is finite. Then the dimension p_k of the subgroup $\mathfrak{P}_k$ of $\mathfrak{M}_k$ $(k \leqq q)$ is finite and is equal to the dimension of the difference group $\mathfrak{M}_{k+1} - \mathfrak{R}_{k+1}$. Similarly the dimension r_k of the k-th homology group $\mathfrak{H}_k$ $(k \leqq q)$ is finite and is equal to the dimension of the difference group $\mathfrak{R}_k - \mathfrak{P}_k$. Moreover the relations*

$$m_0 = r_0 + p_0, \quad m_k = r_k + p_k + p_{k-1} \qquad (k = 1, \cdots, q) \tag{11.1}$$

hold and the differences $h_k = m_k - r_k$ satisfy the inequalities

$$h_k - h_{k-1} + \cdots + (-1)^k h_0 \geqq 0 \qquad (k = 0, 1, \cdots, q). \tag{11.2}$$

By the use of equation (11.1) it is found that the first member of (11.2) is identical with p_k and is accordingly non-negative, as was to be shown.

COROLLARY. *The number of isolated critical k-sets does not exceed m_k.*

In fact the number of isolated critical k-sets may equal m_k, as will be seen in **14** below.

Consider now an isolated set H of homotopic critical points and let $\mathfrak{M}^*_k$ be the class of all F-homology classes M_k whose critical k-sets are in H. As was seen in Theorem 9.3, the class $\mathfrak{M}^*_k$ is a subgroup of $\mathfrak{M}_k$ that is completely determined by the k-cycles on a neighborhood of H. Let m^*_k be the dimension of $\mathfrak{M}^*_k$. If $m^*_k > 0$ the set H will be said to be a homotopic critical k-set of index k of order m^*_k. The order m^*_k of H may be considered to be the number of critical k-sets in H. The set H may be of index k and index $j \neq k$ at the same time. By the use of Theorem 9.3 one obtains the following:

THEOREM 11.2. *Suppose that the dimension m_k of the k-th F-homology group is finite. The number of isolated homotopic critical sets of index k does not exceed m_k and the sum of their orders is equal to m_k.*

With a suitable interpretation this result is also true when m_k is infinite.

12. Extensions of admissible classes $\mathfrak{F}$. A theory analogous to that given above can be developed when the definition of admissible classes $\mathfrak{F}$ is modified in a number of ways. We shall give some examples and indicate briefly the modifications that have to be made. In the first place the assumption that the set $F = S$ is in $\mathfrak{F}$ is a trivial one. This assumption was used essentially to identify the dimensions of the groups $\mathfrak{N}_k - \mathfrak{P}_k$ and $\mathfrak{H}_k$ in Theorem 11. 1. Similarly the assumption that the null set is in $\mathfrak{F}$ can be discarded if the set $F = 0$ is interpreted as the intersection of all sets in $\mathfrak{F}$. The assumption that the sets in $\mathfrak{F}$ are closed is not always essential. For example, for a continuous function $f(x)$ on S the set determined by inequalities of the form $f(x) \leqq c$ and $f(x) < c$ form a class for which the above theory is readily applicable. The theory so developed in analogous to one recently given by Seifert and Threlfall (XVIII).

A more significant modification is obtained if one replaces the hypothesis that given two sets in $\mathfrak{F}$ one is a subset of the other by the assumption that the sum of two sets in $\mathfrak{F}$ is in $\mathfrak{F}$. In this case we define a *primary subset* of F to be a subset F' of F in $\mathfrak{F}$ for which there exist no proper subset F'' of F in $\mathfrak{F}$ such that $F' + F'' = F$. The results described in **3** and **4** then hold with slight modification in their proofs. The results in **5** are also true although the sets X of type A relative to (u, F, F') can no longer be interpreted as those preserving the minimizing properties of (u, F, F') when the points exterior to a neighborhood of X are disregarded. However a set X of type A can be associated with the minimizing properties of a k-cycle (u, F, F') if one observes that the set Y appearing in the definition of such a set plays a role analogous to that of F'' in Lemma 3. 1. The results described in **6** are not valid without further modification. However, those described in **7** and **8** hold true. In **9** the only modification that needs to be made is in the definition of the sums of F-homology classes. This modification is based on the fact that if (u, F, F') and (v, G, G') are minimizing k-cycles and H is the intersection of all sets in $\mathfrak{F}$ containing $F + G - FG$ as a subset then there is a minimizing k-cycle (w, H, H') with $H' = H(F' + G' + FG)$ such that $u + v \sim w$ on $F + G \bmod F' + G' + FG$. Moreover every k-cycle related to (u, F, F') and (v, F, F') in this manner is homologous to (w, H, H'). If $F = G$, then $(w, H, H') = (0, 0, 0)$. Thus if (u, F, F') is the F-homology class M_k and (v, F, F') is in the F-homology class M'_k, the class containing (w, H, H') will be defined as the sum $M_k + M'_k$ of the classes M_k and M'_k. With this modification the results described in **9** are found to be valid. Similarly the results described in the last two sections also hold, although the proofs need to be modified.

There are many examples of admissible classes of the type just described. For example, the class of all closed sets on S, or the class of all compact sets on S with the set $F = S$ adjoined are of this type. In this case the minimizing k-sets are minimum carriers of k-cycles. Similarly, given a homotopic deformation Δ of S the class of all closed sets F in S such that ΔF is deformed on F under Δ is a class of the type here considered. As a somewhat trivial but suggestive example of this case, let S be the lemniscate in the Euclidean two-dimensional space E defined by the equation $r^2 = \cos 2\theta$. We use the metric of E on S. Let O, A, B be respectively the points whose polar coordinates (r, θ) are $(0, 0)$, $(1, 0)$, $(1, \pi)$. Let D be the class of all homotopic deformations Δ of S on S which do not diminish the distance r of a point (r, θ) from the origin O. Let $\mathfrak{F}$ be the class of all closed subsets F of S such that $\Delta F \subset F$ for every deformation Δ in D. The class $\mathfrak{F}$ consists of the set S, the null set and the sums of the following sets: the points O, A, B; the closed subarcs of S having A or B as the initial point and O as the final end point; the closed subsets of S which have A or B as their initial point but do not pass through the point O. The points O, A, B are the only homotopic critical points. The minimizing k-sets are those for which $\Delta F = F$ for every deformation Δ in D. The points O, A, B and their sums are the minimizing 0-sets and critical 0-sets. The sums of the closed subarcs of S having A or B as their initial point and O as their final end-point are minimizing 1-sets and have the point O as their critical 1-set. Let Λ be the integers modulo 2. Then the connectivities r_k of S are given by the set $r_0 = 1$, $r_1 = 2$, $r_k = 0$ $(k \geqq 2)$ and the numbers m_k, p_k corresponding to those in **11** are $m_0 = 3$, $m_1 = 4$, $m_k = 0$ $(k \geqq 2)$; $p_0 = 2$, $p_1 = 2$, $p_k = 0$ $(k \geqq 2)$. The relations (11.1) and hence also (11.2) are accordingly satisfied.

13. Saddle points of functions. In this section we shall consider a real single valued function $f(x) = f(x_1, \cdots, x_n)$ defined over a neighborhood of the closure S of an open set in Euclidean $(x_1 \cdots x_n)$-space. It will be assumed that a neighborhood N of a point x on the boundary B of S can be represented in a one-to-one manner by continuous functions

$$x_i(\alpha, \beta) = x_i(\alpha_1, \cdots, \alpha_{n-1}, \beta) \qquad (i = 1, \cdots, n)$$

such that a point of N is in S if and only if $\beta \leqq 0$. The boundary points of S in N are given by the points $x_i(\alpha, 0)$. The function $g(\alpha) = f[x(\alpha, 0)]$ will be called the *boundary function.* The results described in this section hold also if S can be represented only locally by a Euclidean coordinate system.

Let x_0 be an interior point of S. The point x_0 will be called a *saddle* *point of* $f(x)$ *of index* k if a neighborhood of x_0 can be represented in a one-to-one manner by continuous functions

$$x_i(y, z) = x_i(y_1, \cdots, y_k, z_1, \cdots, z_{n-k}) \qquad (i = 1, \cdots, n)$$

such that the point x_0 corresponds to the point $(y, z) = (0, 0)$ and such that the function $\phi(y, z) = f[x(y, z)]$ has the following two properties: (1) There exists a constant $r > 0$ such that for each point $(y, z) \neq (y, 0)$ satisfying the relations

$$(13.1) \qquad y_1^2 + \cdots + y_k^2 \leqq r^2, \qquad z_1^2 + \cdots + z_{n-k}^2 \leqq r^2$$

the function $\phi(y, hz)$ is an increasing function of h on the interval $0 \leqq h \leqq 1$; (2) For each $(hy, 0) \neq (0, 0)$ on the set (13.1) the function $\phi(hy, 0)$ is a decreasing function of h. From these two properties it is found that $\phi(0, z) > \phi(0, 0) > \phi(y, 0)$ for points $(0, z) \neq (0, 0)$ and $(y, 0) \neq (0, 0)$ on the set (13.1).

A boundary point of x_0 of S will be called a *saddle point of* $f(x)$ *of index* k if a neighborhood of x_0 can be represented by coordinates (y, z) as described above in such a way that the values (y, z) determine a point of S if and only if $z_{n-k} \leqq 0$ and such that properties (1) and (2) described above hold when $z_{n-k} \leqq 0$.

The admissible class $\mathfrak{F}$ of sets F here used is the one described in Example 1 of **6**, namely, the class of all subsets of S determined by an inequality of the form $f(x) \leqq c$, where c is a constant or $c = \pm\infty$. It will be assumed that the sets $F \neq S$ in $\mathfrak{F}$ are bounded. Since they are closed, they must also be compact and $f(x)$ has a finite lower bound on S. It will be assumed further that the coefficient group Λ is a field.

THEOREM 13.1. *Let* x_0 *be a saddle point of* $f(x)$ *of index* k *and let* F *be the set in* $\mathfrak{F}$ *determined by the inequality* $f(x) \leqq f(x_0)$. *The point* x_0 *is a homotopic critical point relative to* F *and there is a neighborhood of* x_0 *containing no other homotopic critical point. In fact the point* x_0 *forms a homotopic critical set of index* k *and order one and is not of index* $j \neq k$. *The index of a saddle point is uniquely defined.*

In order to prove this result we shall restrict ourselves to a saddle point x_0 that is interior to S. The saddle points on the boundary can be treated similarly by restricting the coordinate z_{n-k} to be of constant sign. Let a neighborhood of the point x_0 under consideration be represented by the coordinates (y, z) described above. Denote by c the minimum of the values $\phi(y, z)$ on the set

$$y_\alpha y_\alpha = 0, \qquad z_\sigma z_\sigma = r^2 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k).$$

Here and elsewhere a repeated subscript will be used to denote summation with respect to that subscript. Choose a constant h such that $\phi(0, 0) < h < c$ and let H be the set in $\mathfrak{F}$ determined by the inequality $f(x) \leqq h$. Select a positive constant $s < r$ such that the relation $\phi(y, z) > h$ holds on the set

$$y_\alpha y_\alpha \leqq s^2, \qquad z_\sigma z_\sigma = r^2,$$

and let N be the neighborhood of x_0 determined by the inequalities

$$y_\alpha y_\alpha < s^2, \qquad z_\sigma z_\sigma < r^2.$$

We shall show now that there exists an F-deformation Δ of H which deforms the set $H - HN$ on $H - HN$ and deforms the set HN on HN into the set determined by the relations

$$(13.2) \qquad y_\alpha y_\alpha < s^2, \qquad z_\sigma = 0 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k).$$

To do so let $h(y)$ be a real valued continuous function of $y = (y_1, \cdots, y_k)$ such that $h(y) = 1$ when $y_\alpha y_\alpha \leqq s^2$, $0 \leqq h(y) \leqq 1$ when $s^2 \leqq y_\alpha y_\alpha \leqq r^2$ and $h(y) = 0$ elsewhere. Let Δ be the F-deformation of H defined by the following two properties: The points determined by the values of (y, z) on (13.1) are replaced by the points determined by the values $(\bar{y}, \bar{z})$ according to the law

$$\bar{y}_\alpha = y_\alpha, \qquad \bar{z}_\sigma = (1 - t)h(y)z_\sigma \qquad (0 \leqq t \leqq 1);$$

all other points of H are held fast. On N the values (y, z) are replaced by the values $(y, (1-t)z)$ which are also on N. Since the function $f[y, (1-t)z]$ is a decreasing function of t for fixed values of $(y, z) \neq (y, 0)$, it follows that there is no homotopic critical point on HN except possibly on the set (13.2). We shall see presently that the point $(y, z) = (0, 0)$ is the only homotopic critical point on N. It should be observed that if $k = 0$ the points of HN are deformed into a single point under Δ. The conclusions of the theorem are therefore immediate when $k = 0$. In the sequel we shall suppose that $k > 0$.

Let P be the set obtained by deleting the point x_0 from the set F described in the theorem. The points of the image ΔP of P under the F-deformation Δ which lie in N are determined by the relations

$$(13.3) \qquad 0 < y_\alpha y_\alpha < s^2, \qquad z_\sigma = 0.$$

Let Δ' be the F-deformation of ΔP for which the points of ΔP that are not in N are held fast and such that a point (y, z) in N and hence on the set (13.5) is replaced by the point $(\bar{y}, \bar{z})$ determined by the relations

$$\bar{y}_\alpha = y_\alpha(1 - t + ts/(y_\beta y_\beta)^{1/2}), \qquad \bar{z}_\sigma = 0 \qquad (0 \leqq t \leqq 1).$$

The deformation Δ of P followed by the deformation Δ' of ΔP defines an F-deformation Δ'' of P which carries the points of PN into the $(k-1)$-sphere

$$(13.4) \qquad y_\alpha y_\alpha = s^2, \qquad z_\sigma = 0 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k).$$

It follows that there is no homotopic critical point on the set PN. The point x_0 is accordingly the only possible critical point in HN.

We shall now prove that the point x_0 forms a critical k-set and hence also a homotopic critical set of index k. To do so consider a j-cycle (u, F, F'), where F is the set described in the theorem. Since the value of $\phi(y, z)$ is less than $\phi(0, 0)$ on the set (13.4) we can suppose that the set (13.4) is on the set F' without disturbing the minimizing properties of (u, F, F'). Under the F-deformation Δ described above the j-chain u is deformed into a j-chain v on ΔF such that $u \sim v$ on $F \bmod F'$. We can accordingly suppose that u is on the set ΔF since we can replace u by v. Let $Y = F - FH$. In view of the definition of critical j-sets we are interested only in determining when the relation $u \not\sim 0$ on $F \bmod F' + F'' + Y$ holds for every proper subset F'' of F in $\mathfrak{F}$. We can accordingly suppose that u is on the closure of the set $\Delta(FN)$, that is, on the set

$$(13.5) \qquad y_\alpha y_\alpha \leqq s^2, \qquad z_\sigma = 0.$$

Under the deformation Δ'' of F' determined by the deformation Δ'' of P the boundary u' of u is deformed into a $(k-1)$-chain on the set (13.4). We can suppose, therefore, that u' is on the set (13.4). If $u' \sim 0$ on the set (13.4) we can suppose, by Lemma 2.3, that $u' = 0$ and that u is a j-cycle on the set (13.5). Since the j-th connectivity $(j > 0)$ of the set (13.5) is zero, we have $u \sim 0$ on this set and hence $(u, F, F') \sim 0$ also, in this case. It follows that the point x_0 cannot form a critical j-set $(j \neq k)$. In order to show that the point x_0 forms a critical k-set we observe that there is a $(k-1)$-cycle v' on the set (13.4) which is non-bounding on this set. This $(k-1)$-cycle bounds a k-chain v which is a k-cycle on $F \bmod F'$. In fact if F'' is any proper subset of F in $\mathfrak{F}$ containing F', then $v \not\sim 0$ on $F \bmod F''$. For otherwise we would have $v \sim 0$ on $\Delta F \bmod \Delta F''$ and hence $v \sim 0$ on $\Delta(FN) \bmod \Delta(F''N)$. Under the deformation described above the set $\Delta(F''N)$ is deformed into the set (13.4). It would follow that we would have $v' \sim 0$ on the set (13.4), contrary to our choice of v'. The k-cycle (v, F, F'') is, therefore, a minimizing k-cycle and x_0 is the critical k-set determined by it. Moreover there is an element λ in the field Λ such that $u' \sim \lambda v'$ on the set (13.4). We have accordingly $(u, F, F') \sim (\lambda v, F, F')$. The count of the point x_0 as a critical k-set is, therefore, unity. This completes the proof of Theorem 13.1.

THEOREM 13.2. *Suppose that the homotopic critical points of* $f(x)$ *are all saddle points and let* m_k *be the number of saddle points of* $f(x)$ *of index* k *on* S. *If the numbers* m_k $(k = 0, 1, \cdots, n)$ *are finite, then the connectivities* $r_0, \cdots, r_n$ *of* S *are finite and the numbers* $h_k = m_k - r_k$ *satisfy the relations* (2.3) *and*

$$h_n - h_{n-1} + h_{n-2} - \cdots + (-1)^n h_0 = 0. \tag{13.6}$$

If S *is bounded and closed and if no point of* S *is a limit point of homotopic critical points, these numbers are finite.*

14. Analytical critical points of functions. In the present section it will be assumed that the function $f(x)$ described in the last section has continuous second partial derivatives on a neighborhood of the closed S. It will be assumed further that the functions $x_i(\alpha_1, \cdots, \alpha_{n-1})$ $(i = 1, \cdots, n)$ locally defining the boundary of S have continuous second derivatives and that the matrix of their first derivatives has rank $n - 1$. The boundary function $g(\alpha) = f[x(\alpha)]$ has continuous second derivatives.

A point x_0 at which the first partial derivatives of $f(x)$ are all zero will be called an analytic critical point of $f(x)$. Similarly the analytic critical points of the boundary function $g(\alpha)$ are those at which the first partial derivatives vanish.

THEOREM 14.1. *Let* x_0 *be a homotopic critical point of* $f(x)$ *on* S. *If* x_0 *is interior to* S, *it is an analytic critical point of* $f(x)$. *If* x_0 *is on the boundary* B *of* S *it is an analytic critical point of the boundary function* g *and the outer normal derivative* f_N *of* $f(x)$ *at* $x = x_0$ *satisfies the inequality* $f_N \leqq 0$, *the equality holding only in case the point* x_0 *is also an analytic critical point of* $f(x)$.

In order to prove the first statement suppose that one of the derivatives, say f_{x_1}, is not zero at $x = x_0$. Replace x_1 by $-x_1$ if necessary, so that $f_{x_1} > 0$ on a neighborhood N of x_0. Let Δ be the F-deformation of N defined by the system

$$y_1 = x_1 - bt, \qquad y_j = x_j \qquad (j = 2, \cdots, n;\ 0 \leqq t \leqq 1)$$

where b is a small positive constant. The function $f(y)$ thus obtained is a decreasing function of t for each point x in N. This contradicts our definition of homotopic critical points. The point x is, accordingly, an analytic critical point of $f(x)$. The second statement in the theorem can be proved similarly.

An analytic critical point x_0 of $f(x)$ will be said to be *non-degenerate* if the determinant $|f_{x_i x_j}|$ $(i, j = 1, \cdots, n)$ is different from zero at $x = x_0$. By the *index* of a non-degenerate analytic critical point x_0 will be meant the negative index of the quadratic form $f_{x_i x_j}\xi_i\xi_j$ $(i, j = 1, \cdots, n)$ at $x = x_0$.

THEOREM 14.2. *A non-degenerate analytic critical point x_0 of $f(x)$ of index k, that is interior to S, is a saddle point of index k.*

For suppose that the point x_0 is the origin and that the coordinate system has been chosen so that at $x = x_0$ we have

$$f_{x_i x_j}\xi_i\xi_j = -\xi_1^2 - \cdots - \xi_k^2 + \xi_{k+1}^2 + \cdots + \xi_n^2. \tag{14.1}$$

The $n - k$ independent equations

$$f_{x_j}(x_1, \cdots, x_k, x_{k+1}, \cdots, x_n) = 0 \qquad (j = k+1, \cdots, n) \tag{14.2}$$

hold at $x = 0$ and hence have unique solutions $x_j(x_1, \cdots, x_k)$ $(j = k+1, \cdots, n)$ which are continuous and have continuous derivatives near the values $(x_1, \cdots, x_k) = (0, \cdots, 0)$. The function $\phi(y, z)$ obtained from $f(x)$ by setting

$$x_\alpha = y_\alpha, x_{k+\sigma} = x_{k+\sigma} = x_{k+\sigma}(y_1, \cdots, y_k) + z_\sigma \quad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k)$$

can be shown to have the properties described in the definition of a saddle point of index k. In the first place we have

$$\phi_{z_\sigma}(y, 0) = f_{x_{k+\sigma}}[y_1, \cdots, y_k, x_{k+1}(y), \cdots, x_n(y)] = 0. \tag{14.3}$$

By the use of Taylor's theorem it is found that

$$\phi(y, hz) = \phi(y, 0) + h^2 \int_0^1 (1 - t)\phi_{z_\sigma z_\tau}(y, tz) z_\sigma z_\tau dt. \tag{14.4}$$

The integrand is positive at a point $(y, z) \neq (y, 0)$ on a set of the form

$$y_\alpha y_\alpha \leqq r^2, \qquad z_\sigma z_\sigma \leqq r^2 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k), \tag{14.5}$$

by virtue of the formula $\phi_{z_\sigma z_\tau} = f_{x_{k+\sigma} x_{k+\tau}}$ and the relation (14.5) with $\xi_1 = \cdots = \xi_k = 0$. The function $\phi(y, hz)$ is accordingly an increasing function of h on the set (14.5). On the other hand the function $\phi(y, 0)$ has the continuous first and second derivatives

$$\phi_{y_\alpha} = f_{x_\alpha} + f_{x_j}x_{j y_\alpha} = f_{x_\alpha}[y_1, \cdots, y_k, x_{k+1}(y), \cdots, x_n(y)],$$

$$\phi_{y_\alpha y_\beta} = f_{x_\alpha x_\beta} + f_{x_\alpha x_j}x_{j y_\beta} \qquad (\alpha, \beta = 1, \cdots, k;\ j = k+1, \cdots, n).$$

Differentiating the members of the equation (14.3) with respect to y_α it is found that

$$f_{x_j x_\alpha} + f_{x_j x_l}x_{l y_\alpha} = 0 \qquad (j, l = k+1, \cdots, n).$$

We have accordingly

$$\phi_{y_\alpha y_\beta} = f_{x_\alpha x_\beta} - f_{x_j x_i} x_{i y_\alpha} x_{j y_\beta}.$$

It follows from this relation and formula (14.1) that the quadratic form $\phi_{y_\alpha y_\beta} \xi_\alpha \xi_\beta$ $(\alpha, \beta = 1, \cdots, k)$ is negative definite at $(y, z) = (0, 0)$. From the formula

$$\phi(hy, 0) = \phi(0, 0) + h^2 \int_0^1 (1 - t) \phi_{y_\alpha y_\beta}(ty, 0) y_\alpha y_\beta \, dt$$

we conclude that the function $\phi(hy, 0)$ is a decreasing function of h for each point $(hy, 0) \neq (0, 0)$ on the set (14.5), provided that r is taken sufficiently small. The point x_0 is therefore a saddle point of $f(x)$ of index k, as was to be proved.

By a similar argument one can establish the following:

THEOREM 14.3. *A non-degenerate analytic critical point of index k of the boundary function $g(\alpha)$, at which the outer normal derivative f_N of $f(x)$ is negative is a saddle point of $f(x)$ of index k.*

Combining Theorems 14.2, 14.3 and 13.2 one obtains the following result, first proved by Morse (III) in a simpler form and later extended to the present theorem by Morse and Van Schaack (IX).

THEOREM 14.4. *Suppose that the analytic critical points of $f(x)$ on S are non-degenerate and interior to S. Suppose further that the analytic critical points of the boundary function $g(\alpha)$, at which the outer normal derivative f_N of $f(x)$ is negative, are also non-degenerate. Let m_k be the number of analytic critical points of $f(x)$ of index k on S plus the number of analytic critical points of $g(\alpha)$ of index k at which $f_N < 0$. If the numbers $m_0, \cdots, m_n$ are finite, so also are the connectivities $r_0, \cdots, r_n$ of S and the numbers $h_k = m_k - r_k$ satisfy the relations* (2.3) *and* (12.6). *If the set S is bounded, these numbers are necessarily finite.*

II. APPLICATIONS TO THE CALCULUS OF VARIATIONS.

In the next two sections an application of the above theory will be made to a simple problem in the Calculus of variations. The present treatment is a brief one but is sufficient to bring out some of the essential processes and ideas that are useful in a more extensive treatment. For example, in order to show that every k-cycle is homologous to a minimizing k-cycle it is sufficient to show that the hypotheses of Theorem 4.1 are satisfied, namely, that the sets $f \leqq c$ are compact. The existence of absolute minima is usually established by imposing conditions on the function under consideration which will imply that the sets $f \leqq c$ are compact. Thus it is seen that the results obtained by the various writers on absolute minima can be applied here so as to obtain criteria for the existence of minimizing k-cycles and minimizing k-sets. Moreover in **16** below it is shown that in the positive regular case critical arcs are extremal arcs. This is done by the use of the theory of short extremals. This theory played an important role in the earlier existence theorems for absolute minima. It was used by Birkhoff in order to establish the existence of extremals of index unity. It plays a basic role in the earlier papers of Morse and in the abstract variational theory recently developed by Morse.

15. The positive quasi-regular case. Let $\mathfrak{X}$ be a bounded closed set of points $x = (x_1, \cdots, x_n)$ in an Euclidean n-space. Denote by $\phi(x, p) = \phi(x_1, \cdots, x_n, p_1, \cdots, p_n)$ a real single valued function defined for all elements $(x, p) \neq (x, 0)$ with x in a neighborhood $\mathfrak{N}$ of $\mathfrak{X}$. It will be assumed that for these values of (x, p) the function $\phi(x, p)$ has continuous derivatives of the first three orders and is positively homogeneous of degree one in p. It will be assumed further that for each set (x, p, q) with x in $\mathfrak{X}$ and $p \neq 0$, $q \neq 0$ one has

$$(15.1) \qquad \phi(x, p) > 0, \qquad E(x, p, q) = \phi(x, q) - q_i \phi_{p_i}(x, p) \geqq 0,$$

where $(i = 1, \cdots, n)$. These relations are the so-called conditions of *positive definiteness* and *positive quasi-regularity*[2] for the integral

$$f(X) = \int_{t_1}^{t_2} \phi[x(t), x'(t)] dt.$$

Here X denotes an absolutely continuous curve

$$(15.2) \qquad x_i(t) \qquad (t_1 \leqq t \leqq t_2;\ i = 1, \cdots, n)$$

on $\mathfrak{N}$. In this section and the following one the argument of the function f and hence the points of the space S of the preceding pages will be denoted by capital letters instead of small letters.

Let S^* be the class of all directed rectifiable curves X in $\mathfrak{X}$. Curves consisting of a single point are admitted. Two curves in S^* will be considered identical if they consist of the same points in the same order. We introduce a Fréchet distance on S^* in the usual manner (VII, p. 202). A metric space

---

[2] See for example, McShane (VII, p. 204). Here references are found to the important works of Tonelli, Graves, and others.

S^* is thus obtained. Denote by S a subspace of S^* that is closed relative to S^* and let $\mathfrak{F}$ be the admissible class of sets F on S determined by inequalities of the form $f(X) \leqq c$, where c is a real constant or $+\infty$.

THEOREM 15.1. *The sets $F \neq S$ in $\mathfrak{F}$ are compact.*

It is well known that the conditions (15.1) imply that given a finite constant c there is a constant N such that if X is a curve in S having $f(X) \leqq c$, then $L(X) \leqq N$, where $L(X)$ is the length of X. Moreover the function $f(X)$ is lower semi-continuous on the subset of curves X in S defined by the inequality $L(X) \leqq N$. These results follow, for example, from the proofs of Theorem 1 and Lemma 3 of McShane (VII, p. 209). Consider now a sequence of curves $\{X_r\}$ in S having $f(X_r) \leqq c$. In view of the relations $L(X_r) \leqq N$, the sequence has a rectifiable accumulation curve X_0, by Hilbert's theorem (cf. VII, p. 210). The curve X_0 is in S since S is closed in S^*. Moreover $f(X_0) \leqq \lim f(X_r) \leqq c$, by virtue of the lower semi-continuity of $f(X)$ on the set of curves X having $L(X) \leqq N$. This proves the theorem.

In view of the last theorem it is seen by Theorems 4.1 and 5.1 that critical k-sets exist whenever modular k-cycles exist that are not homologous to zero. Further properties of critical k-sets are yet to be determined. For example it is not known whether the above hypotheses are sufficient to show that on the critical k-set C corresponding to a non-null minimizing k-set F in $\mathfrak{F}$, the relation $f(X) = c$ holds, where c is the least upper bound of $f(X)$ on F. This result is true in the positive regular case as will be seen in the next section.

16. The positive regular case. In the present section we shall make the additional assumption that at each element $(x, p) \neq (x, 0)$ with x in $\mathfrak{R}$ the inequality

$$\phi_{p_i p_k} h_i h_k > 0 \qquad (i, k = 1, \cdots, n) \tag{16.1}$$

holds for every set $(h) \neq (bp)$. This is the so-called *positive regular case.* In this case extremals exist. We shall show that every homotopic critical curve, and hence that every critical curve of index k, is an extremal arc.

An arc X on $\mathfrak{R}$ that can be represented by functions (15.2) having continuous first and second derivatives with $x'_i x'_i \neq 0$ will be called an *extremal arc* if it satisfies the Euler equations

$$\phi_{x_i} - (d/dt)\phi_{p_i} = 0.$$

An extremal X will be called a minimizing extremal if it affords a minimum to $f(X)$ in the class of all rectifiable curves on $\mathfrak{R}$ joining its end points and is similarly directed. It is well known that the hypotheses made above insure the existence of a constant $e > 0$ such that every pair of points in $\mathfrak{X}$ within an e-distance of each other can be joined in either direction by a unique minimizing extremal on $\mathfrak{R}$. These extremals vary continuously with their end points and the integral $f(X)$ is continuous on this class of extremal arcs. We suppose that a constant e has been so chosen.

In the present section it will be assumed that the subclass S of S^* described in **15** has the following additional property: *Let Y be a subarc of a curve X in S such that the diameter of Y does not exceed the constant e chosen above. The space S contains the arc obtained from X by replacing the subarc Y of X by the unique minimizing extremal joining its end points and similarly directed.* This assumption is essentially an assumption of extremal convexity of our space S.

In setting up the F-deformations described below we shall use a special parametrization of curves due to Whitney and Morse (XV). They have shown that a curve X in S can be defined by points

$$x_i(t, X) \qquad (0 \leqq t \leqq 1;\ i = 1, \cdots, n), \tag{16.2}$$

where the functions $x_i(t, X)$ are continuous in t and X for all t on $0 \leqq t \leqq 1$ and X in S. Moreover for each fixed curve X, not a point curve, the functions (15.2) are not all constants on any subinterval of $0 \leqq t \leqq 1$.

LEMMA 16.1. *Let P be a compact subset of S and suppose that the arcs of S have been parameterized in the manner described above. For every constant $\epsilon > 0$ there is a constant $\delta > 0$ such that if $t' \leqq t \leqq t''$ is a subinterval of $0 \leqq t \leqq 1$ of length at most δ, the diameter of the subarc of a curve X in P determined by the subinterval $t' \leqq t \leqq t''$ cannot exceed ϵ.*

This lemma follows readily from the fact that the functions $x_i(t, X)$ in (16.2) are continuous on the compact set of elements (t, X) with t on $0 \leqq t \leqq 1$ and X on P and hence are uniformly continuous on this set.

The results describe below are obtained by the use of two F-deformations $\Delta(t', t'')$ and Δ of a compact set P on S. To define the first of these, consider a compact set P of S and suppose that the arcs of P have the special parametrization described above. Let e be the constant chosen at the beginning of this section and choose δ related to $\epsilon = e$ as described in Lemma 15.1. Let $t' \leqq t \leqq t''$ be a subinterval of $0 \leqq t \leqq 1$ of length at most δ. Denote by $Y(\tau, X)$ $(t' \leqq \tau \leqq t'')$ the curve obtained from a curve X in P by replacing the subarc of X determined by $t' \leqq t \leqq \tau$ by the minimizing extremal joining its end points and similarly directed. We have $Y(t', X) = X$. Since these short minimizing extremals vary continuously with their end points, the function $Y(\tau, X)$ is continuous on S for all values (τ, X) with τ on $t' \leqq \tau \leqq t''$ and X on P. The function $Y(\tau, X)$ $(t' \leqq \tau \leqq t'')$ therefore defines a homotopic deformation $\Delta(t', t'')$ of P on S.

LEMMA 16.2. *The deformation $\Delta(t', t'')$ of a compact set P on S is an F-deformation of P on S. Under this deformation the subarc of a curve X in P determined by the interval $t' \leqq t \leqq t''$ is replaced by a minimizing extremal arc while the remaining points of X are held fast.*

By the use of the deformation $\Delta(t', t'')$ one can define an F-deformation Δ of a compact subset P of S as folliows: Let q be an integer such that $q\delta > 1$, where δ is the constant used in the definition of $\Delta(t', t'')$. Divide the interval $0 \leqq t \leqq 1$ into q equal subintervals by values $t_0 = 0 < t_1 < t_2 < \cdots < t_q = 1$. We then have $0 < t_j - t_{j-1} < \delta$ $(j = 1, \cdots, q)$. The deformation obtained by applying the F-deformations $\Delta(t_{j-1}, t_j)$ $(j = 1, \cdots, q)$ simultaneously to the arcs of P will be called the deformation Δ of P on S. Under this deformation the subarc of X determined by $t_{j-1}t_j$ is replaced by an extremal segment. The curve X is therefore deformed into a broken extremal with at most $q - 1$ corners. On this class of broken extremals the integral $f(X)$ is continuous. Hence we have:

LEMMA 16.3. *Under the deformation Δ the compact set P is deformed into a compact set ΔP on which the integral $f(X)$ is continuous.*

The least upper bound b of $f(X)$ on a minimizing k-set F will be called a *critical value of index k corresponding to F*. The set F is accordingly determined by the inequality $f(X) \leqq b$. As a first result we have:

THEOREM 16.1. *The set $F = S$ is not a minimizing k-set, that is, the critical values of index k which exist are finite.*

For consider a k-cycle (u, F, F') with $F = S$ and let (u_n) be a sequence of vertex k-chains defining u. Let w be a $(k+1)$-chain defined by a sequence (w_n) of vertex $(k+1)$-chains such that $u_n - u_{n+1} - w'_n$ is on F'. The vertices of the chains in these sequences together with their limit points form a compact set P. Under the F-deformation Δ of P described above, the set P is deformed onto a set ΔP on which f is continuous and hence bounded. There is, accordingly, a proper subset G of F in $\mathfrak{F}$ such that $\Delta P \subset G$. Under the deformation Δ the k-chain u is deformed into a k-chain v on G. It follows that $u \sim 0$ on $F = S$ mod G. The k-cycle (u, F, F') is, therefore, not a minimizing k-cycle. This proves the theorem.

THEOREM 16. 2. *Let F be a non-null minimizing k-set and let b be the least upper bound of $f(X)$ on F, that is, the critical value of index k corresponding to F. On the critical k-set C corresponding to F, $f(X) = b$.*

This result follows at once from Corollary 2 of Theorem 7. 2 and the properties of the F-deformation Δ of F described in Lemma 16. 3.

THEOREM 16. 3. *Every homotopic critical curve corresponding to a set $F \neq S$ in $\mathfrak{F}$ and hence every curve in a critical k-set is an extremal arc.*

For let X be a critical curve corresponding to a set $F \neq S$ in $\mathfrak{F}$. The curve X is therefore a fixed curve under the F-deformation $\Delta(t', t'')$ of $P = F$ described above. The subarc of X, determined by an arbitrary interval $t' \leqq t \leqq t''$ of $0 \leqq t \leqq 1$ of length at most δ, is accordingly an extremal arc. It follows that X must also be an extremal, as was to be proved.

THEOREM 16. 4. *Let S be the class of all rectifiable curves X on $\mathfrak{X}$ joining two fixed points and let X_0 be an extremal in S on the interior of $\mathfrak{X}$. Suppose that the end points of X_0 are not conjugate and denote by k the number of points on X conjugate to its initial end point, each conjugate point being counted a number of times equal to its order. The curve X_0 forms a critical k-set whose count is unity and is not a critical curve of index $j \neq k$. For each set $F \neq S$ in $\mathfrak{F}$ containing X_0 there is a neighborhood N of X_0 relative to F containing no other homotopic critical curve.*

The proof of this result will be omitted. It can be established by argument like that used by Morse (XVII, p. 448) in the proof of a similar theorem.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILLINOIS.

REFERENCES.

I. H. Poincaré, "Sur les courbes définies par les équations différentielles," *Journal de Mathématiques*, 4th series, vol. 1 (1885), pp. 167-244.

II. G. D. Birkoff, "Dynamical systems with two degrees of freedom," *Transactions of the American Mathematical Society*, vol. 18 (1917), pp. 199-300.

III. M. Morse, "Relations between the critical points of a real function of n independent variables," *Transactions of the American Mathematical Society*, vol. 27 (1925), pp. 345-396.

IV. S. Lefschetz, "Continuous transformations of manifolds," *Proceedings of the National Academy of Sciences*, vol. 11 (1925), pp. 290-292.

V. G. D. Birkhoff, *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9 (1927).

VI. L. Vietoris, "Ueber den höheren Zusammenhang kompakter Raume und eine Klasse von zusammenhangstreuen Abbildungen," *Mathematische Annalen*, vol. 97 (1927), pp. 454-472.

VII. E. J. McShane, "Semicontinuity in the calculus of variations and absolute minima for isoperimetric problems," *Contributions to the Calculus of Variations*, 1930, University of Chicago Press.

VIII. L. Lusternik and L. Schnirelmann, "Méthodes topologiques dans les problèmes variationnels," *Actualités Scientifiques et industrielles*, vol. 188 (1934),

IX. M. Morse and Van Schaack, "The critical point theory under general boundary conditions," *Annals of Mathematics*, vol. 35 (1934), pp. 545-571.

X. M. Morse, *The Calculus of Variations in the Large*, American Mathematical Society Colloquium Publications, vol. 18 (1934).

XI. G. D. Birkhoff and M. R. Hestenes, "Generalized minimax principle in the calculus of variations," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 96-99, *Duke Mathematical Journal*, vol. 1 (1935), pp. 413-432.

XII. M. Morse and van Schaack, "Abstract critical sets," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 258-262.

XIII. S. Lefschetz, "On critical sets," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 220-222, *Duke Mathematical Journal*, vol. 1 (1935), pp. 392-412.

XIV. M. R. Hestenes, "Minimax principle for functions," *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 115-119.

XV. M. Morse, "A special parametrization of curves," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 915-922.

XVI. M. Morse, *Analysis in the Large.* Lectures delivered at the Institute of Advanced Study 1936-7, Princeton University Press.

XVII. M. Morse, "Functional topology and abstract variational theory," *Annals of Mathematics*, vol. 38 (1937), pp. 386-449.

XVIII. H. Seifert and W. Threlfall, *Variationsrechnung im Grossen*, Hamburger Mathematische Einzelschriften, vol. 24 (1938), pp. 1-115.

XIX. M. Morse, *Functional topology and abstract variational theory*, Mémorial des Sciences Mathématiques, vol. 92 (1939), pp. 1-78.

XX. M. Morse, "Rank and span in functional topology," *Annals of Mathematics*, vol. 41 (1940), pp. 419-454.

XXI. W. Mayer, "A new approach to the critical value theory," *Bulletin of the American Mathematical Society*, vol. 46 (1940), pp. 838-847.

XXII. G. Ewing and M. Morse, "The variational theory in the large including the non-regular case—First paper," *Annals of Mathematics*, vol. 44 (1943), pp. 339-353.

XXIII. G. Ewing and M. Morse, "The variational theory in the large including the non-regular case—Second paper," *Annals of Mathematics*, vol. 44 (1943), pp. 354-374.

THE FOUR-VERTEX THEOREM FOR SURFACES OF CONSTANT CURVATURE.*

By S. B. Jackson.

1. Introduction. The Four-Vertex Theorem, first published by Mukhopadhyaya in 1909 [1], states that every plane oval of class C'' possesses at least four extrema of the curvature, where an oval may be defined as a simple closed curve with non-vanishing curvature. This theorem has been extended to other classes of plane curves by Graustein [2], Fog [6], and the writer [4], and to certain limited classes of space curves [3, 7, 8]. In 1940 the writer extended the theorem in a slightly different direction by considering the extrema of the geodesic curvature on spherical curves [5]. The fundamental result of this paper was that every simple closed spherical curve of class C''', not a circle, has at least four extrema of the geodesic curvature. Such extrema are called geodesic vertices. The aim of the present discussion is to extend the results of this earlier paper to curves on any surface of constant Gaussian curvature. Specifically it will be shown (Theorem 4.1) that every simple closed curve of class C''', not a geodesic circle, in a simply connected region of a surface of constant curvature has at least four geodesic vertices.

The technique of the paper is to establish the existence of a transformation, called a transformation of type I, which maps the given simply connected region of the surface into the plane in such a way that the geodesic vertices of the given curve are carried into the vertices of the corresponding plane curve. The desired result then follows readily from the properties of this transformation and the known facts about the vertices of plane curves [4]. It is shown (**5**) by examples that the requirement that the simple closed curve lie in a simply connected region of the surface is essential.

No four-vertex theorem can be expected for surfaces of variable curvature since it is shown (**6**) that all sufficiently small distance circles [1] about a non-stationary point of the Gaussian curvature have only two geodesic vertices.

---

* Received February 2, 1945; Presented to the American Mathematical Society, April 27, 1945.

[1] A distance circle is what Blaschke [11, § 72] calls an Entfernungskreis, and is the locus of points at a fixed geodesic distance from a given point. On a surface of variable curvature distance circles are not geodesic circles as defined in this paper (**2**).

One type of stationary point is considered (7) and is characterized in terms of the geodesic polar coordinate system with center at this point.

2. Geodesic vertices and monotone arcs. A geodesic vertex of a curve on a surface, Σ, is a point or an arc of constant geodesic curvature for which the geodesic curvature is a relative extremum with respect to the neighboring arcs on either side. The term "relative extremum" is to be understood in the following sense. If $1/a$ is the geodesic curvature at the vertex and $1/\rho$ the geodesic curvature at an arbitrary point, then either $1/\rho - 1/a \geqq 0$ or $1/\rho - 1/a \leqq 0$ in a neighborhood of the vertex. The equality sign may hold at any number of points on the adjacent arcs but it must not hold at all points of either of them.

An arc for which the geodesic curvature is monotone non-decreasing or monotone non-increasing has no geodesic vertices, and conversely an arc with no geodesic vertices has this property. For brevity such an arc will be called monotone. All arcs and curves in the following discussion will be understood to have continuous geodesic curvature unless the contrary is stated.

The curves which have constant geodesic curvature are called geodesic circles.[2] We shall seek to characterize vertices and monotone arcs according to their behavior with respect to their osculating geodesic circles. It will be convenient to refer to points as lying to the left or right of a given directed curve. In such a statement it is understood that the surface is viewed from the tip of the directed normal to Σ and that the surface trihedral forms a right-handed system. Thus, in a coordinate system on Σ, the v-curves are always directed toward the left of the u-curves.

THEOREM 2. 1. *Let C_1 and C_2 be curves of a surface Σ which are tangent in the same direction at p_0, and consider the correspondence established between the curves by the family of geodesics orthogonal to C_1. If, for all corresponding points sufficiently near p_0, $1/\rho_2 \geqq 1/\rho_1$, where $1/\rho_i$ is the geodesic curvature of C_i, then, in some neighborhood of p_0, C_2 lies on or to the left of C_1. The two curves either meet only at p_0 or coincide along a single arc, which necessarily contains p_0.*

Consider an orthogonal coordinate system on Σ with C_1 as the curve $u = 0$, and the geodesics orthogonal to C_1 as the u-curves. If u denotes arc length along these geodesics and v denotes arc length along C_1 measured in the positive direction from p_0, the first fundamental form is

$$ds^2 = du^2 + G\, dv^2,$$

---

[2] These are what Blaschke [11, § 72] calls Krümmungskreise as distinguished from Entfernungskreise. Geodesics are thus special geodesic circles.

where $G(0,v)=1$. Since C_2 is tangent to C_1 at p_0, its equation can be written in the form $u=u(v)$, or equivalently, $\phi(u,v)\equiv u-u(v)=0$, where $u(0)=u'(0)=0$. The geodesic curvature of C_2 is readily found, by Bonnet's formula,[3] to be

$$(2.1)\quad 1/\rho_2=[-2Gu''+2G_u u'^2+G_v u'+GG_u]/[2\sqrt{G}(u'^2+G)^{3/2}].$$

The geodesic curvature of C_1, whose equation is $u=0$, is

$$(2.2)\qquad 1/\rho_1=\bar{G}_u/2\bar{G},$$

where the bar denotes that the functions are evaluated for $u=0$. The hypothesis of the theorem is that for sufficiently small v, $1/\rho_2\geqq 1/\rho_1$. The result of substituting (2.1) and (2.2) into this relation may be written in the form

$$u''\leqq G_u u'^2/G+G_v u'/2G+G_u/2-(u'^2+G)^{3/2}\bar{G}_u/2\bar{G}\sqrt{G}.$$

If the right hand side of the above inequality, which is formally a function of u since it contains $G(u,v)$ and its derivatives, is expanded in a finite Taylor series of two terms, the relation assumes the following form

$$u''\leqq \bar{G}_v u'/2\bar{G}+[2u'^2/\bar{G}+1-(1+u'^2/\bar{G})^{3/2}]\bar{G}_u/2+Au$$

where the function A is bounded. The expression $(1+u'^2/\bar{G})^{3/2}$ may be expanded in a convergent binomial series since, for small values of v, $|u'^2/\bar{G}|<1$. If this is done the first two terms above have a common factor u'. By inserting absolute values on the right hand side, the inequality above thus implies that

$$(2.3)\qquad u''\leqq K|u'|+L|u|,$$

where K and L are sufficiently large positive constants.

Recently R. P. Boas, Jr. has established the following result [9]. Let $f(x)$ be a real function possessing a second derivative, with an absolutely continuous first derivative, and satisfying the following two conditions.

i. $f(0)=f'(0)=0$

ii. $f''(x)\leqq\phi(x)\,|f'(x)|+x^{-1}\psi(x)\,|f(x)|$ in the interval $[0,a]$
where $\phi(x)$ and $\psi(x)$ are positive and integrable. Then either $f(x)\equiv 0$ in some interval $[0,b]$ or $f(x)<0$ in the interval $(0,\operatorname{Min}(a,c)]$ where $\int_0^c(\phi+\psi)\,dx=1$.

---

[3] An inspection of the derivation of this formula [10, pp. 157-158] shows that the ambiguous sign is to be taken negative in this case.

This result applies immediately to the function $u(v)$ if we take $\phi = K$ and $\psi = Lu$. Thus either $u \equiv 0$ in some interval $0 \leqq v \leqq b$ or $u < 0$ in $0 < v \leqq b$, whence C_2 either coincides with C_1 in $[0, b]$ or lies to the left of it in $(0, b]$, with p_0 as the only common point for non-negative values of v.

For negative values of v, reverse the direction on the two curves, similarly reversing the coordinate system. This will change the sign of both curvatures, whence in this direction $1/\rho_1 \geqq 1/\rho_2$ or, equivalently, $-1/\rho_2 \geqq -1/\rho_1$. Substitution as above yields the inequality $-u'' \leqq K \mid u' \mid + L \mid u \mid$, from which we conclude as before that $-u \leqq 0$ or $u \geqq 0$. Thus when it is directed in this way C_2 lies on or to the right of C_1. As above the arcs from p_0 in this direction either meet only at p_0 or coincide along a single arc. This completes the proof of the theorem, since, as originally directed, C_2 lies on or to the left of C_1 in some neighborhood of p_0, and the two curves meet only at p_0 or coincide along a single arc which contains p_0 either as an endpoint or an interior point.

COROLLARY 2. 1. 1. *In the neighborhood of a geodesic vertex, a curve lies entirely to the left or entirely to the right of the osculating geodesic circle according as the geodesic curvature at the vertex is a minimum or a maximum.*

COROLLARY 2. 1. 2. *At an interior point of a monotone arc (or at an arc of constant geodesic curvature interior to a monotone arc) the arc crosses its osculating geodesic circle from right to left or from left to right according as the geodesic curvature is monotone non-decreasing or monotone non-increasing.*

Both corollaries are immediate consequences of the theorem. For Corollary 2. 1. 1 it is only necessary to take the curve and its osculating geodesic circle as the curves C_i. For Corollary 2. 1. 2, let one of the curves, C_i, be the path traced by a point which follows the arc as it approaches the vertex and follows the geodesic circle as it leaves it, while the other curve is the path obtained if the point follows the geodesic circle as it approaches the vertex and the arc as it leaves it.

LEMMA 2. 1. *A necessary and sufficient condition that an arc be monotone is that it cross every osculating geodesic circle at the point (or arc) of contact, the geodesic curvature being monotone non-decreasing or non-increasing according as the crossing is from right to left or from left to right.*

The necessity is precisely Corollary 2. 1. 2. The sufficiency follows from the fact that if the arc were not monotone there would be a vertex, which is impossible by Corollary 2. 1. 1. By exclusion we obtain at once the following characterization of geodesic vertices.

LEMMA 2. 2. *A point (or an arc of constant geodesic curvature) of a curve in the neighborhood of which the curve lies entirely on one side of the osculating geodesic circle is either a vertex or a limit point of vertices. If it is a vertex, the geodesic curvature is a maximum or a minimum according as the curve lies to the right or the left of the geodesic circle.*

It may be noted that if a point is a limit point of vertices there are infinitely many vertices of both kinds in any neighborhood of it.

3. Transformations of type *I*. A transformation of class C''' of a surface Σ to a surface Σ' is said to be of type I if it (a) is locally one-to-one, (b) preserves sense, and (c) carries geodesic circles into geodesic circles.

THEOREM 3. 1. *A transformation of type I carries arcs of monotone non-increasing (non-decreasing) geodesic curvature into arcs of monotone non-increasing (non-decreasing) geodesic curvature. It carries geodesic vertices into geodesic vertices of the same kind or into limit points of such vertices.*

This theorem is an immediate consequence of the definition of transformations of type I and the characterizations of monotone arcs and geodesic vertices contained in Lemma 2. 1, Lemma 2. 2, and Corollary 2. 1. 1.

In this discussion we shall be concerned entirely with the case where Σ' is the Gaussian plane, and this will be understood without further statement in the following development.

LEMMA 3. 1. *Every sufficiently small neighborhood of a point on a surface of constant Gaussian curvature, K, admits a one-to-one transformation of type I.*

If $K > 0$, a sufficiently small neighborhood of any point is applicable to a sphere of the same curvature, which in turn can be mapped into the plane by inversion. This transformation is of type I if the plane is suitably oriented. If $K = 0$, a neighborhood of any point is applicable to the plane, and this transformation is surely of type I. If, finally, $K < 0$, the familiar mapping of the neighborhood of a point into the Poincaré half plane is of type I [11, pp. 156-161], and the lemma is proved.

A neighborhood, on a surface of constant curvature, K, which admits a one-to-one transformation of type I into the plane will be termed a "special neighborhood." Throughout the remainder of the paper Σ will denote a surface of constant curvature unless the contrary is stated.

LEMMA 3. 2. *The one-to-one transformations of type I of a region R* *in a special neighborhood of* Σ *are obtained as the product of any particular one and the general direct circular transformation of the plane,* $w = (az + b)/(cz + d)$, $(ad - bc \neq 0)$.

Such transformations are all clearly of type *I*. Suppose now that T_2 is any one-to-one transformation of type *I* of *R*, and T_1 is a second such transformation. Then $T_2T_1^{-1}$ is a transformation of a region of the plane into a region of the plane which carries circles and lines into circles and lines and which preserves sense. It is well known that the only such transformations are the direct circular transformations. Thus $T_2T_1^{-1} = F$ and $T_2 = FT_1$, where F is a direct circular transformation, and the proof is complete.

LEMMA 3. 3. *A one-to-one transformation of type I of a region R in a special neighborhood of* Σ *is uniquely determined by three pairs of corresponding points.*

This is an immediate consequence of Lemma 3. 2 and the fact that a direct circular transformation is a complex linear transformation and is thus uniquely determined by three pairs of corresponding points.

LEMMA 3. 4. *If two transformations of type I (not necessarily one-to-one) on a region R of* Σ *coincide on a neighborhood of R, they coincide for all points of R.*

For otherwise there would be a point *p* of *R*, every neighborhood of which would have infinitely many points where the transformations coincide, and also points where they do not. By taking a special neighborhood of this point where both transformations are one-to-one, a contradiction to Lemma 3. 3 is obtained.

LEMMA 3. 5. *The transformations of type I (if any) of a region R on* Σ *are the products of any one such transformation and the general direct circular transformation.*

For in a special neighborhood where two transformations are one-to-one, one transformation is the product of the other by a direct circular transformation by Lemma 3. 2, whence by Lemma 3. 4 the relation is true for all points of *R*.

LEMMA 3. 6. *All the transformations of type I of any region R of a special neighborhood N are one-to-one.*

For a transformation of type I on R is one-to-one in a sufficiently small neighborhood of any point p. By Lemma 3. 3, there is a one-to-one transformation of type I for all N which coincides with the given transformation near p. By Lemma 3. 4 the transformations then coincide on R and thus the given transformation is one-to-one on R.

Lemma 3. 7. *If R_1 and R_2 are two regions in a special neighborhood N of Σ, and if R_1 is contained in R_2, a one-to-one transformation of type I on R_1 uniquely determines a transformation of type I on R_2 which must also be one-to-one.*

Three points from R_1 together with their plane images uniquely determine a one-to-one transformation of type I for all of N, which coincides with the given transformation on R_1 by Lemma 3. 3. Thus there is defined on R_2 a one-to-one transformation of type I which coincides with the given one on R_1. The uniqueness of the transformation follows from Lemma 3. 4.

Lemma 3. 8. *Let S' be any closed set interior to a simply connected region S of a surface. Then, for any positive number ϵ, S' can be covered by a finite sequence, $C_1, \cdots, C_m$ of simply connected sub-regions satisfying the following conditions.*

(i) *The diameter of every C_i is less than ϵ.*

(ii) $\sum_{i=1}^{i=k} C_i$ *is simply connected,* $1 \leqq k \leqq m$.

(iii) *The intersection of C_{k+1} with* $\sum_{i=1}^{i=k} C_i$ *is simply connected for* $1 \leqq k \leqq m - 1$, *and is contained in the union of a suitably chosen intersecting pair of the preceding C_i.*

Since S is simply connected, it is, by definition, the topological image of an open square, R; i. e. the correspondence is one-to-one and continuous in both directions. If R' is the set corresponding to S', then R' is closed and thus is contained interior to some closed square R'' in R. Since R'' is closed, the transformation of R'' into S is uniformly continuous, so that for any positive ϵ there exists a $\delta(\epsilon)$ such that if two points of R'' have a distance less than δ their images in S have distance less than ϵ.

Let R'' be divided into a network of squares by equally spaced lines parallel to the sides so that the diagonals of the small squares are less than $\delta/2$, and let the vertices of the net which are interior to R'' be ordered starting at the upper left hand corner and taking in order the points of the first row, then the second row and so on. Let these points be $p_1, \cdots, p_m$. Corresponding to each p_i consider the open square C'_i whose center is p_i and which consists of the four small squares of the net meeting at p_i. These squares all have diagonals less than δ. This sequence of squares C'_i clearly satisfies conditions (ii) and (iii) above, for the regions obtained by adding the squares in order are all surely simply connected, and the intersection of any square with the previous ones consists of its intersection with the square above and the square on its left or either alone. Let C_i be the image of C'_i under the transformation. Conditions (ii) and (iii) are satisfied by the C_i since these properties are topological, while condition (i) is satisfied by construction. This completes the proof of the lemma.

THEOREM 3.2. *There exists a transformation of type I, not necessarily one-to-one, taking any simply connected region S of a surface of constant curvature* Σ *into a region of the plane.*

Consider a sequence of closed regions S_n $(n = 1, 2, \cdots)$, contained in S such that $S_{i+1} \supset S_i$ and $\sum_{i=1}^{\infty} S_i = S$. The images of a sequence of closed squares like R'' of Lemma 3.8 would serve the purpose. Let S_k be any one of these regions. Every point of S_k has a special neighborhood, and by the Heine-Borel Theorem a finite number of these neighborhoods N_i $(i = 1, \cdots, r)$, cover S_k. For any point p of S_k let $\rho(p)$ be half the radius of the largest distance circle[1] about p contained entirely in the closure of some N_i. The function $\rho(p)$ is continuous on the closed set S_k, and thus takes on its smallest value which is some positive number b. Thus the circle of radius b about any point of S_k is a special neighborhood.

Let $C_1, \cdots, C_m$ be the finite sequence of regions whose existence was established in Lemma 3.8, where ϵ is taken as b and the closed set S' as S_k. Any two intersecting C_i are thus contained in a special neighborhood since they are contained in a circle of radius b about any point of their intersection. Since C_1 is a region of a special neighborhood, it has a one-to-one mapping of type I into the plane. We shall show that this mapping can be extended successively to all the other C_i in such a way as to be one-to-one on each C_i. The proof is by induction. Suppose the mapping has been extended to $\sum_{i=1}^{i=r} C_i$, $1 \leqq r < m$ and consider C_{r+1}. The intersection of C_{r+1} with $\sum_{i=1}^{i=r} C_i$, by Lemma 3.8, is connected and is contained in $C_j \cup C_k$, where C_j and C_k are a suitably chosen intersecting pair of the preceding C_i. It may happen that $C_j = C_k$. Since $C_j \cup C_k$ is contained in a special neighborhood, the given mapping must be one-to-one on $C_j \cup C_k$ by Lemma 3.6. The transformation of type I is therefore one-to-one on the region common to C_{r+1} and $\sum_{i=1}^{i=r} C_i$. Inasmuch as C_{r+1} is contained in a special neighborhood, Lemma 3.7 shows that the mapping of type I can be extended to all of C_{r+1} and is one-to-one on C_{r+1}. Thus the mapping is extended to $\sum_{i=1}^{i=r+1} C_i$ and the induction is complete. There exists therefore a transformation of type I on any S_k. It will be noted that in this discussion the terms transformation and mapping have been used interchangeably.

Let T_1 be any transformation of type I on S_1. We shall show that this transformation can be extended to all S_n. It has just been shown that there is a transformation T_2 of type I on S_2. Since T_1 and T_2 are both transformations of type I on S_1, it follows from Lemma 3.5 that there is a direct circular transformation F such that $T_1 = FT_2$ on S_1. The transformation FT_2 is thus a transformation of type I on S_2 which coincides with the given transformation on S_1. The given transformation has thus been extended to S_2. Proceeding by induction, the transformation can now be extended to all S_k. But since every point of S is in some S_k, the transformation can be extended to all points of S and the proof is complete.

4. Transformation of type *I* of a closed curve on Σ.

Let C be a closed differentiable curve in a simply connected region S of Σ, and let S be mapped into the plane by a transformation of type I as is possible by Theorem 3.2. Let S' denote the closed set of points consisting of C and the simply connected regions of S bounded by C, and let $\bar{S}$ be the set of points in the plane into which S' is carried by the transformation. In a sufficiently small neighborhood of any point p of S' the transformation of S' to the corresponding points of $\bar{S}$ is one-to-one. The transformation need not be one-to-one in the large since a point in $\bar{S}$ may have several inverses in S'.

LEMMA 4.1. *A point of $\bar{S}$ has at most a finite number of inverses in S'.*

For if any point $\bar{p}$ in $\bar{S}$ had an infinite number of inverses, then by the Weierstrass-Bolzano Theorem the inverses would have a limit point p which, by continuity, would also be an inverse of $\bar{p}$. But then every neighborhood of p would contain points which map into $\bar{p}$ and the transformation would not be one-to-one near p. Since this is impossible the lemma is proved.

LEMMA 4.2. *In a sufficiently small neighborhood of any point $\bar{p}$ of $\bar{S}$ there are no points with more inverses in S' than $\bar{p}$.*

Let the number of inverses of p be k, which is finite by Lemma 4. 1. If the lemma is false there is a sequence of points $\bar{p}_i$ converging to $\bar{p}$, each of which has at least $k + 1$ inverses. Consider the set of limit points of these inverses. By continuity these are all inverses of $\bar{p}$. About each of the k inverses of $\bar{p}$ let there be taken a neighborhood so small that no two points have the same image in $\bar{S}$. All but a finite number of the inverses of all $\bar{p}_i$ lie in these k neighborhoods. Thus some $\bar{p}_j$ has at least $k + 1$ inverses in these k neighborhoods, whence it has two inverses in some one neighborhood. Since this is false by construction we have a contradiction and the lemma is proved.

LEMMA 4. 3. *If a point $\bar{p}$ of $\bar{S}$ is not the image of a point of C, then there is a neighborhood of $\bar{p}$ every point of which has the same number, k, of inverses. These inverses make up neighborhoods of the k inverses of $\bar{p}$ and the mapping of each of these non-intersecting neighborhoods into the one about $\bar{p}$ is one-to-one and thus topological.*

Let the k inverses of $\bar{p}$ be p_i, $1 \leqq i \leqq k$, none of which lie on C. About each p_i take a neighborhood N'_i which is a special neighborhood, which has no points in common with C so that N'_i is interior to S', and so that no two N'_i have a point in common. The mapping of each N'_i into a neighborhood $\bar{N}'_i$ of $\bar{p}$ is one-to-one. Take a neighborhood $\bar{N}$ of $\bar{p}$ interior to all the $\bar{N}'_i$ and so that all points of $\bar{N}$ have not more than k inverses, as is possible by Lemma 4. 2. Every point of $\bar{N}$ has then exactly k inverses since there is one in each of the non-overlapping N'_i. If N_i denotes the set of points of N'_i which go into $\bar{N}$, these neighborhoods satisfy the conditions of the lemma and the proof is complete.

Let the plane curve which is the image of C be denoted by $\bar{C}$. The curve $\bar{C}$ divides the plane into a certain number of finite regions and one infinite one (it is always possible to choose the mapping so that $\bar{C}$ does not contain the point at infinity itself). Considered in the Gaussian plane the infinite region is simply connected, as are all the finite ones, and in the following discussions we shall refer to them all as the simply connected regions of the plane bounded by $\bar{C}$.

LEMMA 4. 4. *The number of inverses, k, is the same for all points of one of the simply connected regions bounded by the image curve $\bar{C}$.*

By Lemma 4. 3 the number of inverses in such a region is a continuous, integral valued function and is therefore constant over the region. This establishes the result.

Let $\bar{R}$ be any one of the simply connected regions into which $\bar{C}$ divides the plane. The inverse set of $\bar{R}$ in S' is clearly open both in S and S'. Let R be any one of the components of this inverse set. Since on the open set R the transformation is open, its image is an open connected subset of $\bar{R}$, say $\bar{Q}$. Suppose that $\bar{Q}$ has a boundary point $\bar{p}$ interior to $\bar{R}$. Take a neighborhood $\bar{N}$ of $\bar{p}$ interior to $\bar{R}$ which has the property established in Lemma 4. 3. Then since $\bar{N}$ intersects $\bar{Q}$, some one of the inverse neighborhoods N intersects R. Thus $N + R$ is an open connected subset of S' mapping into $\bar{R}$. But this is impossible for R was a component of the inverse set and hence already maximal. Thus no boundary point of $\bar{Q}$ is interior to $\bar{R}$. But the only connected subset of $\bar{R}$ bounded entirely by boundary points of $\bar{R}$ is $\bar{R}$ itself, whence $\bar{Q} = \bar{R}$, and the component R maps into the whole of $\bar{R}$.

By Lemma 4. 3 every point $\bar{p}$ of $\bar{R}$ has a neighborhood which is the topological image of each of the k non-overlapping inverse neighborhoods. If $\bar{T}$ is any closed subset of $\bar{R}$, it follows readily from the Heine-Borel Theorem that there exists an ϵ such that the ϵ-neighborhood of every point $\bar{p}$ of $\bar{T}$ has the property stated above. Now let $\bar{q}$ be any point of $\bar{R}$ and q any one of its inverses in R. If a moving point $\bar{p}$ traces out a path in $\bar{R}$ starting from $\bar{q}$, there is defined by continuity a unique path in R starting from q, for by the above discussion whenever the mapping has reached a point $\bar{t}$ of the path it can be extended throughout the ϵ-neighborhood of $\bar{t}$ and thus does not terminate. In order to apply the argument it is only necessary to take $\bar{T}$ as any closed set in $\bar{R}$ containing the path.

Finally we shall show that the transformation of R into $\bar{R}$ is one-to-one. Let $\bar{q}$ be any point of $\bar{R}$ and q_1 and q_2 any two of its inverses in R. Let q_1 and q_2 be joined by a curve in R. The image curve in $\bar{R}$ is therefore closed. Let this closed curve be drawn together to the point $\bar{q}$, which is possible since $\bar{R}$ is simply connected. Each point of this curve will move continuously over a path in $\bar{R}$, and thus this operation will, by the results of the last paragraph, generate a continuous deformation of the curve q_1q_2 into a point. But this is impossible unless $q_1 = q_2$, whence every point of $\bar{R}$ has exactly one inverse in R, and the transformation is one-to-one on R.

The same argument applies to each of the other components of the inverse of $\bar{R}$ in S'. There are clearly k of these components, since each one contains one inverse of each point. Since the transformation of any of these components R on $\bar{R}$ has been proved one-to-one and continuous, it is topological. In view of this topological character of the transformation, we can conclude that the components are all simply connected, like $\bar{R}$. The results of the last few paragraphs may be summarized in the following lemma.

LEMMA 4. 5. *The inverses of any one of the simply connected regions $\bar{R}$*

of the plane bounded by $\bar{C}$ are a finite set of disjoint simply connected regions, each of which is mapped topologically on $\bar{R}$ by the given transformation of type I.

LEMMA 4. 6. *The boundary of any one of the k inverses, R_j, of a simply connected region, $\bar{R}$, of the plane bounded by $\bar{C}$ maps into the boundary of $\bar{R}$.*

Any boundary point, b, of R_j can be approached by a sequence of points of R_j, whose images in $\bar{R}$ converge to the image of b. If this image were interior to $\bar{R}$ the point b would necessarily be an interior point of some other R_i, whence R_i and R_j would overlap which is impossible by Lemma 4. 5. Conversely, every boundary point of $\bar{R}$ can be approached by a sequence of points in $\bar{R}$ whose inverses in R_j have one or more limit points in S'. Any such limit point, by the continuity of the transformation, maps into the given boundary point of $\bar{R}$. If such a limit point were interior to R_j it would map into the interior of $\bar{R}$. Since this is impossible, all these limit points are boundary points of R_j. It should be clearly noted that the mapping on the boundary need not be one-to-one.

An arc AB of a closed plane curve $\bar{C}$ is called a simple loop if (a) it is simple except that $A = B$, and (b) one of the regions into which it divides the plane contains no further points of the curve. This region is then one of the simply connected regions of the plane bounded by $\bar{C}$, and is referred to as the region corresponding to the loop.

Up to this point C has been any closed curve of class C'' in S. Now let C be in particular a simple closed curve, positively directed, i. e. with its interior to its left.

LEMMA 4. 7. *If C is a positively directed simple closed curve of class C'' in the simply connected region S of Σ, then under a transformation of type I the corresponding plane curve $\bar{C}$ can contain no simple loop whose corresponding region lies to its left.*

Suppose that the lemma is false. Then there is a simple loop whose region $\bar{R}$ lies to its left. This region may be assumed finite since in any case it can be made so by a direct circular transformation. Since $\bar{C}$ is the image of C, there is some arc ab of C which maps into the loop, where both a and b map into the double point of the loop. The points close to arc ab and to its left map into the points close to the loop and to its left, i. e. into $\bar{R}$, since the transformation preserves sense. Thus ab is a part of the boundary of one of the inverse regions R_i of $\bar{R}$. Since ab is not closed it cannot be the entire boundary of R_i. Let p be any other point of the boundary of R_i which does not map into the double point of the loop. Such points p exist, since ab and all the inverses of the double point together still do not bound a region in $\bar{S}$. By Lemma 4. 6, p maps into a point $\bar{p}$ of the loop. Take a neighborhood of p in S so small that the mapping in this neighborhood is topological. This neighborhood may include points of S not in S' and the transformed neighborhood at $\bar{p}$ may include points of the plane not in $\bar{S}$. Take a smaller neighborhood $\bar{N}$ if necessary about $\bar{p}$ which is divided in exactly two parts by the loop. This is possible since the loop is differentiable. Since the mapping of $\bar{N}$ on its inverse neighborhood N at p is topological, N is divided into exactly two half neighborhoods by the inverse of the arc. Some point in one of these half neighborhoods is in R_i, and by hypothesis this maps into $\bar{R}$, i. e. to the left of the directed loop at $\bar{p}$. All points of this half neighborhood are thus in R_i since any two can be joined by a path whose image lies entirely in $\bar{R}$. But there is also a point p' of ab which maps into $\bar{p}$, and the set of points to the left of ab near p' also maps into the points of $\bar{R}$ near $\bar{p}$ and to the left of the loop. Thus a point of $\bar{R}$ near $\bar{p}$ has one inverse in R_i near p and a second near p'. This is impossible, since by Lemma 4. 5 the transformation of R_i on $\bar{R}$ is one-to-one. Thus we have a contradiction, and the lemma is proved.

THEOREM 4. 1. *Every simple closed curve of class C'', not a geodesic circle, in a simply connected region of a surface of constant curvature Σ has at least four geodesic vertices.*

Let the simple closed curve C be mapped on the plane by a transformation of type I as discussed above. By Theorem 3. 1 the geodesic vertices of C go into the vertices of the plane curve $\bar{C}$, and by Lemma 4. 7 $\bar{C}$ has no simple loop having its corresponding region to its left. But it is known [4, Theorem 5. 1] that a plane curve having only two vertices has two simple loops, one of which has the corresponding region to its left. Thus $\bar{C}$ cannot have only two vertices, whence $\bar{C}$ and therefore C has at least four geodesic vertices and the proof is complete.

It may be observed that this discussion establishes the Four-Vertex Theorem for a very large number of curves which may, at first glance, appear to be outside the class under consideration. Many plane curves for example are not simple but become simple when viewed as lying on a many sheeted covering surface of the plane, and are thus covered by the above theorem by taking the covering surface for the surface Σ. Similar remarks apply to curves on other surfaces, such as cylinders and cones, which are not simple on these surfaces but are simple on some covering surface.

5. Counterexamples.

This section is devoted to two examples showing that the restriction in Theorem 4.1 that C lie in a simply connected region of Σ cannot be dispensed with.

Example 1.[4] Consider, as the surface Σ, the right circular cylinder with generators parallel to the z-axis and with the circle $x = a\cos\theta$, $y = a\sin\theta$ in the xy-plane as directing curve, and let the curve C be the intersection of Σ with the plane $z = by$. If s denotes the arc length on the directing circle measured from the x-axis, then $s = a\theta$. For the points of C we have therefore $z = by = ba\sin\theta = ba\sin(s/a)$, where $0 \leqq s < 2\pi a$. If the cylinder is rolled out on a plane, C goes into the curve, $y' = ba\sin(x'/a)$, $0 \leqq x' < 2\pi a$. But in one period a sine curve has exactly two vertices, whence the same is true of C. C is thus a simple closed curve with only two geodesic vertices on a surface of constant curvature. It will be noted that C cannot be deformed into a point and thus is not in a simply connected region of Σ. This shows that in general it is not possible to eliminate the requirement that C lie in a simply connected region of Σ, if Theorem 4.1 is to hold.

Example 2. This example will show furthermore that if Theorem 4.1 is to be valid it is not even possible to allow a singularity of the surface in the simply connected region S. Let the surface Σ be a right circular cone with its vertex at $(0, 0, h)$ and the circle $x = a\cos\theta$, $y = a\sin\theta$ in the xy-plane for directing curve, and let C be the intersection of this cone with the plane $z = by$, where $b < h/a$ so that the plane actually cuts all the generators of the lower nappe of the cone. C is then a simple closed curve on Σ with a singular point, namely the vertex, in its interior. If (r, θ, z) are cylindrical coordinates of a general point of C, the following relations are immediate.

$$(5.1) \qquad \begin{aligned} x &= r\cos\theta, \qquad y = r\sin\theta \\ r/a &+ z/h = 1. \end{aligned}$$

Let s have the same meaning as in Example 1, l be the slant height of the cone from the vertex to the xy-plane, ϕ be defined as s/l, and ρ be the distance from the vertex to the general point of C. It is easy to verify that

$$(5.2) \qquad \begin{aligned} l^2 &= a^2 + h^2, \qquad s = a\theta \\ \phi &= (a/l)\theta, \qquad \rho = (l/a)r. \end{aligned}$$

Substitution from equations (5.1) and (5.2) into the relation $z = by$ yields the following relation between ρ and ϕ, for the equation of C:

---

[4] This example was first suggested to the writer by the late W. C. Graustein. It was also suggested by the referee.

$$\rho = \frac{hl}{h + ab \sin (l\phi/a)}, \quad h > ab, \quad 0 \leqq \phi < 2\pi a/l. \tag{5.3}$$

If the cone is cut along the line $\theta = 0$ and flattened out on a plane, the quantities ρ and ϕ of equation (5.3) become precisely the polar coordinates of the points on the corresponding plane curve, and since geodesic curvature is preserved by the operation, it is only necessary to seek the vertices of the plane curve having (5.3) as its polar coordinate equation. Straightforward computation shows that the curvature is

$$K = \frac{[a - bh \sin (l\phi/a)][h + ab \sin (l\phi/a)]^3}{al[h^2 + a^2b^2 + b^2h^2 + 2abh \sin (l\phi/a) - b^2h^2 \sin^2 (l\phi/a)]^{3/2}}. \tag{5.4}$$

Consider now the numerical case $a = 1$, $b = 1/2$, $h = 2$. Equation (5.4) becomes

$$K = \frac{(1 - \sin \sqrt{5}\phi)(4 + \sin \sqrt{5}\phi)^3}{\sqrt{5}(21 + 8 \sin \sqrt{5}\phi - 4 \sin^2 \sqrt{5}\phi)^{3/2}}. \tag{5.5}$$

If $dK/d\phi$ is equated to zero to find the extreme values of K, it appears that ϕ must be a solution of one or the other of the following equations.

$$\begin{aligned} &4 \cos^3 \sqrt{5}\phi - 52 \cos^2 \sqrt{5}\phi - 8 \cos \sqrt{5}\phi - 69 = 0 \\ &\cos \sqrt{5}\phi = 0. \end{aligned} \tag{5.6}$$

The first of equations (5.6), considered as a cubic equation in $\cos \sqrt{5}\phi$, has only one real root and this is bigger than 1, so no value of ϕ will satisfy this equation. The second equation has only two solutions in the interval $0 \leqq \phi < 2\pi/\sqrt{5}$, whence C has only two geodesic vertices. Thus the conclusion of Theorem 4.1 is invalid for the curve C.

It is of some interest to note, in connection with Example 2, that not all the curves C obtained in the manner described have just two vertices. For example the curve C obtained by taking $a = 1$, $b = 1/2$, $h = 1$ turns out to have six geodesic vertices, as may be readily verified by carrying out the computation as indicated above.

6. Surfaces of variable curvature. In this section it will be shown that Theorem 4.1 cannot be extended to surfaces of variable Gaussian curvature. Specifically we shall show that on such surfaces sufficiently small distance circles [1] have only two geodesic vertices.

Let o be any point of a surface, Σ, of variable Gaussian curvature, and let (u, v) be geodesic polar coordinates on Σ with origin at o. The geodesic curvature of the v-curves, that is, of the distance circles about o, is given [10, p. 149] by

$$1/\rho = (1/\sqrt{G})(\partial\sqrt{G}/\partial u). \tag{6.1}$$

We seek the zeros of $\partial(1/\rho)/\partial v$ for $u = \text{const.} \neq 0$, as v varies from 0 to 2π, for the geodesic vertices of this distance circle are precisely the points where $\partial(1/\rho)/\partial v$ changes sign. After differentiation (6.1) becomes

$$(6.2) \qquad \frac{\partial}{\partial v}\left(\frac{1}{\rho}\right) = \frac{1}{\sqrt{G}}\,\frac{\partial^2\sqrt{G}}{\partial u\partial v} - \frac{1}{G}\,\frac{\partial\sqrt{G}}{\partial u}\,\frac{\partial\sqrt{G}}{\partial v}.$$

Since $G = 0$ only at $u = 0$, it suffices to find the zeros of

$$(6.3) \qquad G\,\frac{\partial}{\partial v}\left(\frac{1}{\rho}\right) = \sqrt{G}\,\frac{\partial^2\sqrt{G}}{\partial u\partial v} - \frac{\partial\sqrt{G}}{\partial u}\,\frac{\partial\sqrt{G}}{\partial v}.$$

Assuming sufficient differentiability for Σ, we may expand $\sqrt{G}$ in a power series in u. It is known [10] that $(\sqrt{G})_{(0,v)} = 0$, $(\partial\sqrt{G}/\partial u)_{(0,v)} = 1$. The other coefficients in the expansion may be computed from these by using the Gauss equation, $\partial^2\sqrt{G}/\partial u^2 = -K\sqrt{G}$, and its derivatives at $u = 0$. The subscripts above indicate evaluation for $u = 0$ but v arbitrary. By this method we obtain

$$(6.4) \qquad \sqrt{G} = u - \tfrac{1}{6}K_{(0,v)}u^3 - \tfrac{1}{12}(\partial K/\partial u)_{(0,v)}u^4 + \tfrac{1}{120}(K^2 - 3\partial^2K/\partial u^2)_{(0,v)}u^5 + \cdots.$$

Differentiation of (6.4) and substitution in (6.3) yields the result

$$G\partial(1/\rho)/\partial v = -\tfrac{1}{4}(\partial^2K/\partial u\partial v)_{(0,v)}u^4 - \tfrac{1}{10}(\partial^3K/\partial u^2\partial v)_{(0,v)}u^5 + \cdots.$$

Since, for our purpose, $u \neq 0$, it is enough to find the zeros of $\phi(u, v)$, where

$$(6.5) \qquad \phi(u, v) = \tfrac{1}{4}(\partial^2K/\partial u\partial v)_{(0,v)} + \tfrac{1}{10}(\partial^3K/\partial u^2\partial v)_{(0,v)}u + \cdots$$

for small fixed u and variable v.

Let C_1 and C_2 be two perpendicular curves on the surface Σ meeting at o, with arc lengths s_1 and s_2 respectively, and let θ be the angle from C_1 to a curve C through o. Precisely as in the case of the plane, the directional derivative of a function F on Σ in the direction of C at o is given by

$$\partial F/\partial s = (\partial F/\partial s_1)\cos\theta + (\partial F/\partial s_2)\sin\theta$$

where $\partial F/\partial s_i$ is the directional derivative of F in the direction of C_i. We shall apply this to the Gaussian curvature K on Σ. Let C_1 and C_2 be the curves $v = 0$ and $v = \pi/2$, respectively. Then $\partial F/\partial s_1 = (\partial K/\partial u)_{(0,0)}$, $\partial F/\partial s_2 = (\partial K/\partial u)_{(0,\pi/2)}$, $\partial F/\partial s = (\partial K/\partial u)_{(0,v)}$, $\theta = v$, and the above equation becomes

$$(6.6) \qquad (\partial K/\partial u)_{(0,v)} = (\partial K/\partial u)_{(0,0)}\cos v + (\partial K/\partial u)_{(0,\pi/2)}\sin v.$$

By differentiating (6.6) the following relation is obtained.

$$(6.7)\quad (\partial^2K/\partial u\partial v)_{(0,v)} = -(\partial K/\partial u)_{(0,0)}\sin v + (\partial K/\partial u)_{(0,\pi/2)}\cos v.$$

If $(\partial^2K/\partial u\partial v)_{(0,v)} \equiv 0$, then by (6.7) $(\partial K/\partial u)_{(0,0)} = (\partial K/\partial u)_{(0,\pi/2)} = 0$, whence, by (6.6), $(\partial K/\partial u)_{(0,v)} \equiv 0$, and o is a stationary point of K on Σ. Let o now be a non-stationary point of K. Such points exist, for otherwise Σ would have constant curvature. There are then exactly two values of v on the interval $0 \leqq v < 2\pi$ for which $(\partial^2K/\partial u\partial v)_{(0,v)} = 0$, namely the solutions of the equation

$$(6.8)\quad \tan v = \frac{(\partial K/\partial u)_{(0,\pi/2)}}{(\partial K/\partial u)_{(0,0)}}.$$

These solutions may be denoted by v_1 and v_2 where $v_2 = v_1 + \pi$.

Consider first a point $(0,\bar{v})$, where $\bar{v} \neq v_1, v_2$. Then $\phi(0,\bar{v}) = (1/4)(\partial^2K/\partial u\partial v)_{(0,\bar{v})} \neq 0$. By continuity there exists a neighborhood of the coordinates $(0,\bar{v})$ for which $\phi(u,v) \neq 0$.[5] On the other hand consider the points $(0,v_i)$, $i = 1, 2$. By definition of v_i (6.8) holds, whence it follows that

$$(6.9)\quad \sin v_i = \pm\frac{(\partial K/\partial u)_{(0,\pi/2)}}{m},\qquad \cos v_i = \pm\frac{(\partial K/\partial u)_{(0,0)}}{m}$$

where $m^2 = (\partial K/\partial u)^2_{(0,0)} + (\partial K/\partial u)^2_{(0,\pi/2)} \neq 0$, since o is not a stationary point of K. Differentiation of (6.7) with respect to v and substitution from (6.9) yields the result

$$\begin{aligned}(6.10)\quad (\partial^3K/\partial u\partial v^2)_{(0,v)} &= -(\partial K/\partial u)_{(0,0)}\cos v - (\partial K/\partial u)_{(0,\pi/2)}\sin v\\ &= \mp m(\cos v\cos v_i + \sin v\sin v_i) = \mp m\cos(v - v_i).\end{aligned}$$

Hence $(\partial^3K/\partial u\partial v^2)_{(0,v_i)} = \mp m \neq 0$, and

$$(6.11)\quad \partial\phi(0,v_i)/\partial v \neq 0.$$

By (6.11) and the implicit function theorem, there exists a neighborhood of the coordinates $(0,v_i)$[5] in which the equation $\phi(u,v) = 0$ can be solved for v uniquely and exhaustively. Hence to every point of the set $(0,v)$, $0 \leqq v \leqq 2\pi$, has been ordered a neighborhood of this set of coordinates. By the Heine-Borel Theorem there exists a finite set of these neighborhoods covering this set. A positive u_0 can then be chosen so small that (u_0,v) is always in one of the neighborhoods selected. For such a u_0 there are only two zeros of $\phi(u_0, v)$. Hence $u = u_0$ has only two geodesic vertices. We have proved the following theorem.

---

[5] This is a neighborhood of the coordinate pair $(0,\bar{v})$, i.e., the set of pairs (u,v) such that $|u| < \delta$, $|v - \bar{v}| < \delta$. It is not a neighborhood of o on Σ.

THEOREM 6. 1. *The distance circles of sufficiently small radius about a non-stationary point of the Gaussian curvature on a surface Σ have only two geodesic vertices.*

This result makes it clear that no four-vertex theorem can be expected on surfaces of variable Gaussian curvature.

Theorem 6. 1 makes it possible to establish the following characterization of surfaces of constant curvature.

THEOREM 6. 2. *A necessary and sufficient condition that a surface Σ be of constant curvature is that some neighborhood of every point admit a transformation of type I into the plane.*

The necessity is implied by Theorem 3. 2. It remains to prove the sufficiency. Let p be a non-stationary point of the Gaussian curvature, and suppose a region containing p admits a transformation of type I. Since the transformation is locally one-to-one, a sufficiently small distance circle about p will be carried into a simple closed plane curve, which therefore has four vertices or is a circle. But by Theorem 3. 1 this means that the distance circle has four vertices or none contrary to Theorem 6. 1. This is a contradiction, whence every point is a point of stationary Gaussian curvature, and the surface has constant curvature.

7. Stationary points of the Gaussian curvature. Among the stationary points of the curvature mentioned in **6**, one type will be given further analysis. Let us consider a point o of a surface Σ for which all sufficiently small distance circles have constant geodesic curvature. Such points must have stationary Gaussian curvature, as was seen in **6**. A vertex of an ellipsoid of revolution, for example, is such a point.

Using the coordinate system of **6**, the condition that the distance circles about o have constant geodesic curvature is expressed by the equation $\partial(1/\rho)/\partial v \equiv 0$ for all v and all sufficiently small u, where $1/\rho$ is the geodesic curvature of the distance circles. By (6. 1) this is equivalent to

$$\frac{\partial}{\partial v}\,\frac{1}{\sqrt{G}}\,\frac{\partial\sqrt{G}}{\partial u} = 0 \tag{7. 1}$$

whence there follows

$$\frac{1}{\sqrt{G}}\frac{\partial\sqrt{G}}{\partial u} = 1/u + a(u) \tag{7.2}$$

where $a(u)$ has no singularity near $u = 0$. The form of the function on the right side follows from (6.4). By integrating (7.2) we obtain the following equation

$$\sqrt{G} = b(v)\exp\{\textstyle\int[1/u + a(u)]du\} = ub(v)\exp[A(u)] \tag{7.3}$$

where $A(u) = \int a(u)du$. As in **6**, $(\partial\sqrt{G}/\partial u)_{(0,v)} = 1$, whence it appears that $b(v) = \exp[-A(0)]$. From this there follows

$$\sqrt{G} = u\exp[A(u) - A(0)]; \tag{7.4}$$

that is, G is a function of u alone. Moreover condition (7.4) is sufficient as well as necessary, for a $\sqrt{G}$ of form (7.4) satisfies the conditions for geodesic polar coordinates and certainly satisfies (7.1). We have thus established the following theorem.

THEOREM 7.1. *A necessary and sufficient condition that all distance circles of sufficiently small radius about a point o of a surface Σ have constant geodesic curvature is that the third coefficient, G, of the first fundamental form for the geodesic polar coordinate system about o be a function only of the geodesic distance from o.*

COROLLARY 7.1.1. *A necessary and sufficient condition that all distance circles of sufficiently small radius about a point o of a surface Σ have constant geodesic curvature is that these distance circles and the geodesics through o form an isometric system.*

The corollary follows at once from Theorem 7.1 and the conditions that the parametric curves form an isometric system [10, p. 133], namely $F = 0$, and $E/G = U(u)/V(v)$.

UNIVERSITY OF MARYLAND.

BIBLIOGRAPHY.

---

1. S. Mukhopadhyaya, "New methods in the geometry of a plane arc," *Bulletin of the Calcutta Mathematical Society*, vol. 1 (1909), pp. 31-37.

2. W. C. Graustein, "Extensions of the four-vertex theorem," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 9-23.

3. W. C. Graustein and S. B. Jackson, "The four-vertex theorem for a certain type of space curves," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 737-741.

4. S. B. Jackson, "Vertices of plane curves," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 564-578.

5. ———, "The four-vertex theorem for spherical curves," *American Journal of Mathematics*, vol. 62 (1940), pp. 795-812.

6. D. Fog, "Über den Vierscheitelsatz und seine Verallgemeinerungen," *Sitzungsberichte der Berlin Akademie der Wissenschaft* (1933), pp. 251-254.

7. W. Süss, "Ein Vierscheitelsatz bei geschlossenen Raumkurven," *The Tôhoku Mathematical Journal*, vol. 29 (1928), pp. 359-362.

8. T. Takasu, "Vierscheitelsatz für Raumkurven," *The Tôhoku Mathematical Journal*, vol. 39 (1934), pp. 292-298. There are also several other papers by the same author on this general subject.

9. R. P. Boas, Jr., "A differential inequality," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 95-96.

10. W. C. Graustein, *Differential Geometry*, Macmillan (1935).

11. W. Blaschke, *Vorlesungen über Differentialgeometrie* I, Springer (1930).

CONFORMAL CLASSES OF SURFACES.*

By AARON FIALKOW

Introduction. In previous papers[1] [2, 3, 4], we have developed a method for the study of those properties of a subspace V_n $(1 \leqq n \leqq m-1)$ imbedded in any enveloping space V_m which remain unchanged under conformal transformations of V_m. The pictorial interpretation of the analytic apparatus described in these papers will appear elsewhere. The present paper applies some of these general rsults to a discussion of various classes of surfaces V_2 in Euclidean 3-space R_3 which have a simple conformal character. The results also apply to conformally equivalent surfaces in any conformally Euclidean 3-space. The reader is referred to [4] for the notation used as well as for a statement of the assumptions concerning the class of the functions which occur in this paper.

2. **The analytic apparatus.** The results stated below are a summary of some of the analytic aspects, used in this paper, of the conformal theory of a hypersurface V_n $(n > 1)$ in R_{n+1}. The details of this theory for the general case of a V_n in any V_m and the proofs appear in [4]. Let $a_{\alpha\beta}$, $\left\{ {\alpha \atop \beta\gamma} \right\}$ be the metric tensor and Christoffel symbols of the second kind of a Euclidean space R_{n+1} in some coördinate system $\{y^\alpha\}$. Let

$$(2.1) \qquad y^\alpha = y^\alpha(x^1, x^2, \cdots, x^n)$$

be the equations of a V_n imbedded in R_{n+1}. Let g_{ij}, $\left\{ {i \atop jk} \right\}$ be the induced metric tensor and the Christoffel symbols of the second kind of V_n in the parameter system $\{x^i\}$. Covariant differentiation of (2.1) with respect to $\left\{ {\alpha \atop \beta\gamma} \right\}$, $\left\{ {i \atop jk} \right\}$, denoted by a comma, gives $y_{,ij}{}^\alpha = b_{ij}\zeta^\alpha$, where the b_{ij} are the coefficients of the second fundamental form of V_n and ζ^α is the unit normal to V_n. The mean curvature normal μ^α of V_n is defined by

$$(2.2) \qquad \mu^\alpha = 1/n \cdot y_{,ij}{}^\alpha g^{ij}.$$

Its length is the mean curvature μ so that

---

* Received June 20, 1944.

[1] The numbers refer to the bibliography at the end of the paper.

$$\mu = (1/n) b_{ij} g^{ij} = (1/n) \sum_{i=1}^{n} {}_{(i)}b.$$

The ${}_{(i)}b$ are the n principal normal curvatures of V_n.

Let $\{\bar{y}^\alpha\}$ be another coördinate system in R_{n+1} and suppose that the point transformation $\bar{y}^\alpha = y^\alpha$ is a conformal mapping of R_{n+1} on itself. Since the differential elements of distance at corresponding points of R_{n+1} are related by the equation

$$d\bar{s} = e^\sigma ds \tag{2.3}$$

it follows that $\bar{a}_{\alpha\beta} = e^{2\sigma} a_{\alpha\beta}$. The conformal mapping $\bar{V}_n \leftrightarrow V_n$ induced by the parametric equations $\bar{x}^i = x^i$ is such that $\bar{g}_{ij} = e^{2\sigma} g_{ij}$. The tensor $G_{ij}{}^\alpha$ defined by

$$G_{ij}{}^\alpha = y_{,ij}{}^\alpha - g_{ij}\mu^\alpha \tag{2.4}$$

is unchanged by conformal transformations of R_{n+1}. The *relative conformal curvature* e^ϕ is defined by

$$e^{2\phi} = (1/n) a_{\alpha\beta} g^{hi} g^{jk} G_{hj}{}^\alpha G_{ik}{}^\beta = (1/n) \sum_{i=1}^{n} ({}_{(i)}b - \mu)^2. \tag{2.5}$$

It is the standard deviation of the statistical population consisting of the n principal normal curvatures ${}_{(1)}b, {}_{(2)}b, \cdots, {}_{(n)}b$. The *gauge function* ϕ has the transformation law

$$\bar{\phi} = \phi - \sigma. \tag{2.6}$$

The *gauge vector* η^α given by

$$\eta^\alpha = \phi_{,i} g^{ij} y_{,j}{}^\alpha + \mu^\alpha \tag{2.7}$$

has the transformation law

$$\bar{\eta}_\alpha = \eta_\alpha - \sigma_{,\alpha} \tag{2.8}$$

relative to conformal maps of R_{n+1}. The *conformal measure tensors* $A_{\alpha\beta}, G_{ij}$ and the *conformal coefficients of connection* $\Gamma_{\beta\gamma}{}^\alpha, \Gamma_{jk}{}^i$ are defined by

$$A_{\alpha\beta} = e^{2\phi} a_{\alpha\beta}, \qquad G_{ij} = e^{2\phi} g_{ij} \tag{2.9}$$

$$\begin{aligned} \Gamma_{\beta\gamma}{}^\alpha &= \left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\} + \delta_\beta{}^\alpha \eta_\gamma + \delta_\gamma{}^\alpha \eta_\beta - a_{\beta\gamma} a^{\alpha\delta} \eta_\delta \\ \Gamma_{jk}{}^i &= \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} + \delta_j{}^i \phi_{,k} + \delta_k{}^i \phi_{,j} - g_{jk} g^{ih} \phi_{,h} \end{aligned} \tag{2.10}$$

and remain unchanged by conformal transformations of R_{n+1}. The $\Gamma_{jk}{}^i$ are the Christoffel symbols of the second kind formed with respect to G_{ij}. The δ's in (2.10) are the Kronecker deltas.

Conformal length or *conformal distance* S is measured by means of $A_{\alpha\beta}$, G_{ij}; that is, for a displacement dy^{α} from a point of V_n,

$$dS^2 = A_{\alpha\beta} dy^{\alpha} dy^{\beta}. \tag{2.11}$$

If dy^{α} lies in the tangent vector space of V_n and its surface components are dx^i,

$$dS^2 = G_{ij} dx^i dx^j. \tag{2.12}$$

It is clear that

$$dS = e^{\phi} ds \tag{2.13}$$

where ds is the element of distance of R_{n+1} defined by $ds^2 = a_{\alpha\beta} dy^{\alpha} dy^{\beta}$ or on V_n, by $ds^2 = g_{ij} dx^i dx^j$.

Conformal differentiation, denoted by a colon, is defined as covariant tensor differentiation relative to the conformal coefficients of connection. For example,

$$\begin{aligned} T^{\alpha i}_{\beta j:k} = \partial T_{\beta j}{}^{\alpha i}/\partial x^k + T_{\beta j}{}^{\gamma i} \Gamma_{\gamma\delta}{}^{\alpha} y_{,k}{}^{\delta} + T_{\beta j}{}^{\alpha h} \Gamma_{hk}{}^{i} \\ - T_{\gamma j}{}^{\alpha i} \Gamma_{\beta\delta}{}^{\gamma} y_{,k}{}^{\delta} - T_{\beta h}{}^{\alpha i} \Gamma_{jk}{}^{h}. \end{aligned} \tag{2.14}$$

The conformal derivative of a conformal tensor is a conformal tensor. Other tensors defined on V_n which are invariant relative to conformal mappings of R_{n+1} are:

Conformal normal measure tensor B_{ij}

$$B_{ij} = e^{\phi} [b_{ij} - \mu g_{ij}]. \tag{2.15}$$

Deviation tensor E_{ij}

$$\begin{aligned} E_{ij} &= \mu_{\alpha} G_{ij}{}^{\alpha} - \phi_{,ij} + \phi_{,i}\phi_{,j} + \tfrac{1}{2}[\mu^2 - g^{hk}\phi_{,h}\phi_{,k}] g_{ij} \\ &= \mu_{\alpha} G_{ij}{}^{\alpha} - \phi_{:ij} - \phi_{:i}\phi_{:j} + \tfrac{1}{2}[e^{-2\phi}\mu^2 + G^{hk}\phi_{:h}\phi_{:k}] G_{ij}. \end{aligned} \tag{2.16}$$

Deviation vector F_i

$$F_i = (1/(n-1)) \tag{2.17}$$

(Surface) conformal Riemann tensor $K_{jkl}{}^{i}$

$$K_{jkl}{}^{i} = (\partial/\partial x^k) T_{jl}{}^{i} - (\partial/\partial x^l) \Gamma_{jk}{}^{i} + \Gamma_{jl}{}^{h} \Gamma_{hk}{}^{i} - \Gamma_{jk}{}^{h} \Gamma_{hl}{}^{i}. \tag{2.18}$$

The tensors G_{ij}, B_{ij}, E_{ij} are called the *first, second and third conformal fundamental tensors* of V_n, respectively, and serve as the coefficients of the *conformal fundamental forms* of V_n. The tensor B_{ij} satisfies the algebraic relations

$$B_{ij} G^{ij} = 0, \qquad B_{ij} B_{kl} G^{ik} G^{jl} = n. \tag{2.19}$$

Differential conditions on the conformal tensors are

$$(2.20) \qquad B_{ik}B_{hj} - B_{ij}B_{hk} = K_{hijk} + G_{ij}E_{hk} - G_{ik}E_{hj} + G_{hk}E_{ij} - G_{hj}E_{ik}$$

$$(2.21) \qquad B_{ij:k} - B_{ik:j} = G_{ik}F_j - G_{ij}F_k$$

$$(2.22) \qquad F_{j:k} - K_{k:j} = B_{hk}E_j^h - B_{hj}E_k^h.$$

$$(2.23) \qquad E_{ij:k} - E_{ik:j} = B_{ij}F_k - B_{ik}F_j.$$

The basic theorems are:

(1) *Equivalence theorem*: If the conformal fundamental forms of two hypersurfaces V_n and $\bar{V}_n$ are equal, then a conformal transformation of R_{n+1} on itself exists so that V_n is mapped on $\bar{V}_n$.

(2) *Existence theorem*: A hypersurface in R_{n+1} always exists having prescribed conformal fundamental forms, provided the tensor coefficients of these forms satisfy equations (2.19) to (2.23).

3. The case $n=2$. In surface theory $n=2$, $m=3$. In this case, we may choose the lines of curvature of V_2 as parametric. We write the surface parameters x^1, x^2 as u, v, respectively, and the components of the conformal measure tensor as

$$(3.1) \qquad \begin{array}{lll} G_{11} = X & G_{12} = 0 & G_{22} = Y \\ G^{11} = X^{-1} & G^{12} = 0 & G^{22} = Y^{-1}. \end{array}$$

According to (2.15), the principal directions determined by B_{ij} relative to G_{ij} are the lines of curvature. (This fact establishes the conformal invariance of the lines of curvature.) It follows from (2.19) that the components of B_{ij} are

$$(3.2) \qquad B_{11} = X \qquad B_{12} = 0 \qquad B_{22} = -Y.$$

The Christoffel symbols Γ_{jk}^i defined by (2.10) become

$$(3.3) \qquad \begin{array}{lll} \Gamma_{11}^1 = X_u/2X & \Gamma_{12}^1 = X_v/2X & \Gamma_{22}^1 = -Y_u/2X \\ \Gamma_{11}^2 = -X_v/2Y & \Gamma_{12}^2 = Y_u/2Y & \Gamma_{22}^2 = Y_v/2Y. \end{array}$$

From (3.2) and (3.3),

$$(3.4) \qquad \begin{array}{ll} B_{11:1} = B_{11:2} = & B_{22:1} = B_{22:2} = 0 \\ B_{12:1} = -X_v & B_{12:2} = Y_u. \end{array}$$

Also, from (2.17), (3.1), (3.3) and (3.4),

$$
\begin{aligned}
F_1 &= [\log Y]_u & F_2 &= -[\log X]_v \\
F_{1:2} &= [\log Y]_{uv} & F_{2:1} &= -[\log X]_{uv}.
\end{aligned} \tag{3.5}
$$

Using the above equations, it is easy to show that equations (2.21) are true identically when $n = 2$. Equations (2.20) are equivalent to

$$K = 2 - 2E \tag{3.6}$$

where K and E are defined by

$$K = K_{jki}{}^{i} G^{jk} = 1/2(XY)^2 \times [2XY(X_{vv} + Y_{uu}) - Y_u(XY)_u - X_v(XY)_v] \tag{3.7}$$

$$E = E_{ij} G^{ij}. \tag{3.8}$$

Since the *conformal Gaussian curvature* J of V_2 is defined by

$$J = K_{1212}/XY \tag{3.9}$$

it follows from (3.7) and (3.9) that

$$K = -2J \tag{3.10}$$

so that (3.6) may be written as

$$E = J + 1. \tag{3.11}$$

Equation (2.22) becomes

$$[\log \sqrt{XY}]_{uv} + E_{12} = 0 \tag{3.12}$$

and (2.23) becomes

$$\partial E_{11}/\partial v = \partial E_{12}/\partial u + \tfrac{1}{2}E_{12}\,[\log Y/X]_u + \tfrac{1}{2}X_v(E - 2) \tag{3.13}$$

$$\partial E_{22}/\partial v = \partial E_{12}/\partial v - \tfrac{1}{2}E_{12}\,[\log Y/X]_v + \tfrac{1}{2}Y_u(E - 2). \tag{3.14}$$

If ${}_{(1)}L^i$, ${}_{(2)}L^i$ are the unit conformal tangent vectors to the lines of curvature, then, according to (3.1) and (3.2), G_{ij} and B_{ij} have the canonical representation

$$
\begin{aligned}
G_{ij} &= {}_{(1)}L_i\,{}_{(1)}L_j + {}_{(2)}L_i\,{}_{(2)}L_j \\
B_{ij} &= {}_{(1)}L_i\,{}_{(1)}L_j - {}_{(2)}L_i\,{}_{(2)}L_j.
\end{aligned} \tag{3.15}
$$

It follows from the known theory [1, pp. 107-113] that any symmetric second order tensor T_{ij} whose principal directions are tangent to the lines of curvature must have the canonical representation

$$T_{ij} = {}_{(1)}a\,{}_{(1)}L_i\,{}_{(1)}L_j + {}_{(2)}a\,{}_{(2)}L_i\,{}_{(2)}L_j \tag{3.16}$$

where ${}_{(1)}a$ and ${}_{(2)}a$ are the principal invariants of T_{ij}. In virtue of (3.15), (3.16) implies

$$T_{ij} = \tfrac{1}{2}[({}_{(1)}a + {}_{(2)}a)G_{ij} + ({}_{(1)}a - {}_{(2)}a)B_{ij}] \tag{3.17}$$

and conversely. We also note that

$$\mu = \tfrac{1}{2}[{}_{(1)}b + {}_{(2)}b], \qquad e^{\phi} = \tfrac{1}{2}\,|\,{}_{(1)}b - {}_{(2)}b\,|. \tag{3.18}$$

In view of our choice of parameters u, v,

$${}_{(1)}L_i = X^{1/2}\delta_i^{\,1} \qquad {}_{(2)}L_i = Y^{1/2}\delta_i^{\,2}$$

where the δ's are the Kronecker deltas. The *conformal* (*surface*) *principal normal vectors* ${}_{(1)}N_i$, ${}_{(2)}N_i$ of the lines of curvature are defined as

$$\begin{aligned} {}_{(1)}N_i &= d\,{}_{(1)}L_i/d\,{}_{(1)}S - \Gamma_{ij}^{\;h}\,{}_{(1)}L_h\,{}_{(1)}L^j = -\,{}_{(1)}N \cdot {}_{(2)}L_i \\ {}_{(2)}N_i &= d\,{}_{(2)}L_i/d\,{}_{(2)}S - \Gamma_{ij}^{\;h}\,{}_{(2)}L_h\,{}_{(2)}L^j = -\,{}_{(2)}N\,{}_{(1)}L_i \end{aligned} \tag{3.19}$$

where $d\,{}_{(1)}S$ and $d\,{}_{(2)}S$ indicates directional differentiation with respect to S in the directions ${}_{(1)}L_i$ and ${}_{(2)}L_i$, respectively, and ${}_{(1)}N$ and ${}_{(2)}N$ are the corresponding *conformal* (*surface*) *curvatures* of the lines of curvature. ${}_{(1)}N$ and ${}_{(2)}N$ are the analogues of "geodesic curvature." From (3.19), we obtain

$${}_{(1)}N = X_v/2XY^{1/2} \qquad {}_{(2)}N = Y_u/2YX^{1/2}. \tag{3.20}$$

From (3.20), it follows that

$$\begin{aligned} d\,{}_{(1)}N/d\,{}_{(1)}S &= 1/2\sqrt{XY}\{[\log X]_{uv} - \tfrac{1}{2}[\log X]_v[\log Y]_u\} \\ d\,{}_{(2)}N/d\,{}_{(2)}S &= 1/2\sqrt{XY}\{[\log Y]_{uv} - \tfrac{1}{2}[\log X]_v[\log Y]_u\}. \end{aligned} \tag{3.21}$$

4. **Gradient surfaces.** The *deviation* of V_2 for the direction determined by the unit conformal vector L^i is $E_{ij}L^iL^j$. Since E_{ij} is a symmetric tensor, it follows from the known theory that its principal directions relative to G_{ij} are two orthogonal directions at each point of V_2. These are the *deviation directions* and the corresponding curves are the *lines of deviation.* The corresponding deviations are the *principal deviations* ${}_{(1)}E$ and ${}_{(2)}E$. They are the roots of the determinantal equation $|E_{ij} - \rho G_{ij}| = 0$. It follows from (3.8) $F_{;ij}\,{}_{(2)}L^i\,{}_{(1)}L^j = 0$.

$$E = {}_{(1)}E + {}_{(2)}E. \tag{4.1}$$

Of course, the deviation directions are conformally invariant. When do the lines of curvature and the lines of deviation of V_2 coincide? Since the lines of curvature are parametric, the necessary and sufficient condition is $E_{12} = 0$. It follows from (3.12) that this condition is equivalent to

$$(4.2)\qquad [\log \sqrt{XY}]_{uv} = 0$$

which leads to

$$X/f^2(u) = g^2(v)/Y = Q(u, v).$$

Consequently, after a parameter transformation

$$(4.3)\qquad u' = \int f(u)\,du, \qquad v' = \int g(v)\,dv$$

which preserves the parametric directions,

$$dS^2 = Q \cdot du'^2 + 1/Q \cdot dv'^2.$$

Dropping the primes, we may suppose

$$(4.4)\qquad X = Q \qquad Y = 1/Q.$$

Conversely, (4.4) implies (4.2) and hence $E_{12} = 0$. Now the conformal area of any curvilinear rectangle bounded by the lines of curvature of parameters u_1, u_2, v_1, v_2 is

$$\int\int \sqrt{XY}\,du\,dv = (u_2 - u_1)(v_2 - v_1).$$

Hence the lines of curvature form a (conformal) equiareal pattern. This proves the following theorem:

THEOREM 4.1. *A necessary and sufficient condition that the lines of curvature and the lines of deviation of a surface coincide is that the lines of curvature be an equiareal net (relative to the element of conformal area).*

According to (3.5), (4.2) is also equivalent to

$$(4.5)\qquad F_{1:2} = F_{2:1}.$$

As is well known, (4.5) is the condition that F_i be a gradient, that is,

$$(4.6)\qquad F_i = F_{:i}$$

where F is a scalar. It follows from (3.5) that F may be defined as $\log Y$ in this case. We call a V_2 for which (4.6) is true a *gradient surface.* The preceding discussion proves

THEOREM 4.2. *The lines of curvature and the lines of deviation of a V_2 coincide if and only if the V_2 is a gradient surface.*

Finally, we note that in accordance with (3.17), if V_2 is a gradient surface,

$$E_{ij} = \tfrac{1}{2}[EG_{ij} + E'B_{ij}] \tag{4.7}$$

where

$$E' = {}_{(1)}E - {}_{(2)}E. \tag{4.8}$$

5. Isothermic surfaces. In classical differential geometry, surfaces whose lines of curvature form an isothermal net are called isothermic. The isothermal character of a net is conformally invariant since such a net is the conformal image of a rectangular net of straight lines. Hence isothermic surfaces remain isothermic under conformal transformations of R_3. We derive a conformal characterization of these surfaces below.

If V_2 is isothermic, it follows from (2.9) and (3.1) that

$$X = Y \tag{5.1}$$

in some parameter system for which the lines of curvature are parametric, and conversely. It follows from (3.5) that

$$F_{1:2} + F_{2:1} = 0. \tag{5.2}$$

Conversely, (5.2) implies (5.1) after parameter transformations of the type (4.3). Equation (5.2) implies that the principal directions of the symmetric tensor $F_{i:j} + F_{j:i}$ relative to G_{ij} are tangent to the lines of curvature. The concluding remarks of **3** leading to (3.17) then prove

THEOREM 5.1. *The tensor* $F_{i:j} + F_{j:i}$ *is a linear combination of the first and second conformal fundamental tensors of* V_2 *if and only if* V_2 *is isothermic.*

We also note that

$$F_{1:2} + F_{2:1} = 2\sqrt{XY}[d_{(2)}N/d_{(2)}S - d_{(1)}N/d_{(1)}S]$$

as follows from (3.5) and (3.21). As a consequence of the preceding theorem, *the equation*

$$d_{(1)}N/d_{(1)}S = d_{(2)}N/d_{(2)}S$$

is a characteristic property of isothermic surfaces.

6. Surfaces of constant conformal Gaussian curvature. A surface for which J is constant is called a *surface of constant conformal Gaussian curvature*. Exactly as in classical differential geometry, such surfaces admit three-parameter groups of point transformations on themselves (*conformal motions*) which preserve the element of conformal length dS. They also enjoy those other properties of surfaces of constant curvature which are derivable from the first fundamental form of the surface. A different kind of property of these surfaces follows directly from (3.11) and (4.1): *The sum of the principal deviations of a* V_2 *is constant if and only if the* V_2 *has constant conformal Gaussian curvature.*

The simplest of these surfaces are those for which J is zero and whose conformal metric, referred to the lines of curvature, is Cartesian. In this case

$$X = Y = 1 \tag{6.1}$$

so that

$$\begin{aligned} G_{ij}dx^i dx^j &= du^2 + dv^2 \\ B_{ij}dx^i dx^j &= du^2 - dv^2. \end{aligned} \tag{6.2}$$

It follows from (3.12) that $E_{12} = 0$ and from (3.13) and (3.14) that

$$E_{11} = E_{11}(u), \qquad E_{22} = E_{22}(v). \tag{6.3}$$

Since (4.1) is true, it follows from (3.11) and (6.3) that

$$E_{11} = c, \qquad E_{12} = 0, \qquad E_{22} = 1 - c \tag{6.4}$$

where c is a constant. Also, according to (3.5) and (6.1),

$$F_1 = 0, \qquad F_2 = 0. \tag{6.5}$$

Conversely, (6.5) implies (6.1) in view of (3.5). Now it is shown in a forthcoming paper that the line of curvature, $v =$ constant, is a circle if and only if $F_1 = 0$ along its length. Hence (6.5) implies that all the lines of curvature of V_2 are circular. Of course, this is a conformally invariant property. It is a characteristic property of Dupin cyclides and proves

THEOREM 6.1. *If the lines of curvature of a* V_2 *form a Euclidean Cartesian net relative to the conformal element of length on* V_2, *then* V_2 *is a Dupin cyclide, and conversely.*

The simplest Dupin cyclides are the right circular cone, right circular cylinder and proper torus. The first and second conformal fundamental forms of each of these surfaces are given by (6.2) after a suitable choice of parameters. Below, we determine the third conformal fundamental form in each case, using the definitions of **2** and **3**. The equations and components of the fundamental metric tensors of a right circular cone are

$$\begin{array}{lll} y^1 = \omega \cos \alpha & y^2 = \omega \sin \alpha & y^3 = a\omega \\ g_{11} = \omega^2 & g_{12} = 0 & g_{22} = 1 + a^2 \\ b_{11} = 0 & b_{12} = 0 & b_{22} = a\omega/(\sqrt{1 + a^2}). \end{array}$$

After the parameter transformation

$$u = a\alpha/(2\sqrt{1 + a^2}) \qquad v = (a/2)\log\omega$$

we find that (6.2) is true and that

$$(6.6) \qquad E_{ij}dx^idx^j = ((3/2) + 2/a^2)du^2 + (-(1/2) - (2/a^2))dv^2.$$

The corresponding equations and tensor components for a right circular cylinder are

$$\begin{array}{lll} y^1 = a\cos\alpha & y^2 = a\sin\alpha & y^3 = \omega \\ g_{11} = a^2 & g_{12} = 0 & g_{22} = 1 \\ b_{11} = 0 & b_{12} = 0 & b_{22} = a. \end{array}$$

After the parameter transformation

$$u = 1\alpha/2 \qquad v = 1\omega/2a$$

we obtain (6.2) and

$$(6.7) \qquad E_{ij}dx^idx^j = (3du^2)/2 - (1dv^2)/2.$$

The corresponding equations and tensor components for a proper torus are

$$\begin{array}{lll} y^1 = (a + d\cos\beta)\cos\alpha & y^2 = (a + d\cos\beta)\sin\alpha & y^3 = d\sin\beta, \quad a > d \\ g_{11} = (a + d\cos\beta)^2 & g_{12} = 0 & g_{22} = d^2 \\ b_{11} = (a + d\cos\beta)\cos\beta & b_{12} = 0 & b_{22} = d. \end{array}$$

After the parameter transformation

$$u = a\alpha/2d \qquad v = (a/2)\int[d\beta/(a + d\cos\beta)]$$

we establish (6.2) and

$$(6.8) \qquad E_{ij}dx^idx^j = (-(1/2) + 2d^2/a^2)du^2 + ((3/2) - (2d^2/a^2))dv^2.$$

In all the above equations, a, d are constants and α, β, ω are variables.

A comparison of these results with (6.3) shows that every possible form $E_{ij}dx^idx^j$ for a Dupin cyclic is actually realized by the cone, cylinder and proper[2] torus. Now all Dupin cyclides have the first and second conformal fundamental forms which appear in (6.2). Therefore, in accordance with the conformal equivalence theorem noted in 2, we have the following theorem:

THEOREM 6.2. *Every Dupin cyclide is conformally equivalent to a right circular cone, right circular cylinder or proper torus, and conversely.*

---

[2] The improper torus for which $a = d$ or $a < d$ is conformally equivalent to the cylinder or cone, respectively, as may be seen by a comparison of $E_{ij}dx^idx^j$ for these surfaces.

As an example of a surface having non-zero constant conformal Gaussian curvature we mention the catenoid whose equations are

$$y^1 = \omega \cos \alpha \qquad y^2 = \omega \sin \alpha \qquad y^3 = a \log (\omega + \sqrt{\omega^2 - a^2}).$$

From these equations we obtain

$$\begin{array}{lll} g_{11} = \omega^2 & g_{12} = 0 & g_{22} = \omega^2/(\omega^2 - a^2) \\ b_{11} = a & b_{12} = 0 & b_{22} = a/(a^2 - \omega^2). \end{array}$$

After a straightforward calculation and the parameter transformation

$$u = \alpha \qquad v = \operatorname{arc\,sec} (\omega/a)$$

we find that

$$\begin{aligned} G_{ij}dx^i dx^j &= \cos^2 v\, du^2 + dv^2 \\ B_{ij}dx^i dx^j &= \cos^2 v\, du^2 - dv^2 \\ E_{ij}dx^i dx^j &= \qquad\qquad 2dv^2. \end{aligned}$$

From the last of these equations and (3.11) and (4.1), it follows that $J = 1$.

We note that, as is stated earlier, if

$$(6.9) \qquad F_1 = 0$$

throughout a V_2, then all the lines of curvature of V_2, $v =$ constant, are circles. Hence (6.9) is a characteristic property of any surface which is the envelope of an arbitrary one parameter family of spheres (if the lines of curvature are parametric). In any parameter system the corresponding condition is

$$(6.10) \qquad F_i\, {}_{(1)}L^i = 0.$$

Since, from (3.15),

$${}_{(1)}L_i\, {}_{(1)}L_j = \tfrac{1}{2}[G_{ij} + B_{ij}]$$

(6.10) may be written as

$$F_i + B_{ij}F^j = 0$$

where $F^j = G^{jk}F_k$.

7. Isothermic gradient surfaces. According to (4.5) and (5.2), isothermic gradient surfaces are characterized by the equations

$$(7.1) \qquad F_{1:2} = F_{2:1} = 0.$$

Since $F_i = F_{;i}$ in this case, we may write the above equation in the form

$$F_{;ij}\, {}_{(1)}L^i\, {}_{(2)}L^j = F_{;ij}\, {}_{(2)}L^i\, {}_{(1)}L^j = 0.$$

This means that the principal directions of $F_{i;j}$ are tangent to the lines of curvature so that *$F_{i;j}$ is a linear combination of G_{ij} and B_{ij}.*

We now determine the distinctive character of the conformal fundamental forms of such surfaces. Since V_2 is a gradient surface, (4.4) is true. The isothermic character of V_2 implies (5.2) which, using (3.5) and (4.4), leads to

$$Q = Q_2(v)/Q_1(u), \tag{7.2}$$

where Q_1 and Q_2 are, respectively, functions of u and v only. Conversely (7.2) implies (7.1) which characterizes isothermic surfaces.

It remains to impose the condition that

$$dS^2 = (Q_2/Q_1)\,du^2 + (Q_1/Q_2)\,dv^2 \tag{7.3}$$

be the first conformal fundamental form of an actual surface. In view of the existence theorem stated in 2, it is necessary and sufficient to impose conditions (3.11) to (3.14). Equation (3.12) is satisfied identically since (4.4), (7.2) and the equation $E_{12} = 0$ are true. According to (3.7), (3.10), (4.4) and (7.2),

$$J = -\tfrac{1}{2}[Q_2''/Q_1 + Q_1''/Q_2] \tag{7.4}$$

where the primes refer to differentiation, with respect to the independent variable, of the functions Q_1, Q_2. From (3.11), (4.1) and (4.8), we obtain

$$2\,{}_{(1)}E = 1 + J + E', \qquad 2\,{}_{(2)}E = 1 + J - E'. \tag{7.5}$$

Also, since

$$E_{11} = {}_{(1)}E(Q_2/Q_1), \qquad E_{22} = {}_{(2)}E(Q_1/Q_2), \qquad E_{12} = 0$$

(3.13) and (3.14) become

$${}_{(1)}E_v = -(Q'_2/Q_1)(E'/2 + 1), \qquad {}_{(2)}E_u = (Q'_1/Q_2)(E'/2 - 1). \tag{7.6}$$

Differentiation of (7.5) yields

$$\begin{aligned} {}_{(1)}E_v &= E_v'/2 - \tfrac{1}{4}(Q_2'''/Q_1 - Q_2'Q_1''/Q_2^2) \\ {}_{(2)}E_u &= -E_u'/2 - \tfrac{1}{4}(Q_1'''/Q_2 - Q_1'Q_2''/Q_1^2) \end{aligned} \tag{7.7}$$

after use is made of (7.4). From (7.6) and (7.7), we obtain

$$\begin{aligned} E_v' &= -(Q'_2/Q_2)(2 + E') + \tfrac{1}{2}(Q_2'''/Q_1 - Q_2'Q_1''/Q_2^2) \\ E_u' &= \quad (Q'_1/Q_1)(2 - E') - \tfrac{1}{2}(Q_1'''/Q_2 - Q'Q_2''/Q_1^2). \end{aligned}$$

The integrability condition of these equations is $E'_{vu} = E'_{uv}$ or

$$\begin{aligned} 8Q_1'Q_2'/Q_1Q_2 + Q_1'Q_2'''/Q_1^2 + Q_2'Q_1'''/Q_2^2 \\ + Q_1'Q_1''Q_2'/Q_1Q_2^2 + Q_2'Q_2''Q_1'/Q_2Q_1^2 = 0. \end{aligned} \tag{7.8}$$

Let us suppose that $Q_1'Q_2' \neq 0$. Then multiplication of (7.8) by $Q_1^2Q_2^2/Q_1'Q_2'$ and successive differentiation with respect to u and then v leads to the equation, $Q_1'Q_2' = 0$, which contradicts the hypothesis. The remaining possibility, $Q_1'Q_2' = 0$, is clearly a solution of (7.8). Without loss of generality (interchanging u and v if necessary) we may assume that

$$Q_1' = 0, \qquad Q_1 = 1, \qquad Q_2 = Q(v).$$

In this case, the form (7.3) may be written as

$$dS^2 = Q^2(v)\,du^2 + dv^2 \tag{7.9}$$

after the parameter transformation

$$u' = u, \qquad v' = \int (dv/\sqrt{Q})$$

and writing $Q^2(v')$ for $Q[v(v')]$ and dropping primes.

We have shown that the first conformal fundamental form of every isothermic gradient surface may be written as (7.9). Conversely, if $Q(v)$ in (7.9) is an arbitrary function of v, the surface is isothermic gradient. Indeed, from (3.11) to (3.14), we find that the only E_{ij} which are possible are the coefficients of the form

$$\begin{aligned} E_{ij}dx^idx^j = {} & [-(1/2)\,Q^2 - (1/2)\,Q'^2 + c]\,du^2 \\ & + [3/2 + \tfrac{1}{2}\,\frac{Q'^2}{Q^2} - Q''/Q - c/Q^2]\,dv^2 \end{aligned} \tag{7.10}$$

where c is a constant. Then the existence theorem of **2** affirms the existence of surfaces for which (7.9) and (7.10) are the first and third conformal fundamental forms. Of course, the second conformal fundamental form is

$$B_{ij}dx^idx^j = Q^2(v)\,du^2 - dv^2. \tag{7.11}$$

It is clear that we may also choose u, v so that the lines of curvature are parametric and

$$dS^2 = Q^2(v)\,[du^2 + dv^2]. \tag{7.12}$$

In this case, the second and third conformal fundamental forms are

$$\begin{aligned} B_{ij}dx^idx^j &= Q^2(v)\,[du^2 - dv^2] \\ E_{ij}dx^idx^j &= [-(1/2)\,Q^2 - \tfrac{1}{2}\,\frac{Q'^2}{Q^2} + c]\,du^2 \\ &\quad + [(3/2)\,Q^2 + \tfrac{3}{2}\,\frac{Q'^2}{Q^2} - Q''/Q - c]\,dv^2 \end{aligned} \tag{7.13}$$

where c is the same constant that occurs in (7.10). We also note, as follows from (3.5) and (7.12), that (6.9) is true. Hence all the lines of curvature, v = constant, are circles. We state some of these facts in the following theorem:

THEOREM 7.1. *If V_2 is an isothermic gradient surface, its first conformal fundamental form is defined by* (7.9) *in some parameter system for which the lines of curvature are parametric, and conversely. The other conformal fundamental forms are then defined by* (7.10) *and* (7.11).

The simplest isothermic gradient surfaces are those for which $Q(v)$ in (7.12) is a constant. In this case, in accordance with Theorem 6.1, the surface is a Dupin cyclide. We now consider the problem of characterizing all isothermic gradient surfaces in an analogous manner.

Let V_2 be an isothermic gradient surface and suppose that its (metric) first and second fundamental forms referred to the parameters u, v of (7.12) are

$$\begin{aligned} g_{ij}dx^i dx^j &= e^{2\alpha}(du^2 + dv^2) \\ b_{ij}dx^i dx^j &= e^{\alpha}(\gamma du^2 + \delta dv^2) \end{aligned} \tag{7.14}$$

where α, γ, δ are functions of u and v. In these parameters, the classical Gauss-Codazzi equations become

$$\gamma\delta + \alpha_{uu} + \alpha_{vv} = 0 \tag{7.15}$$

$$\gamma_v - \delta\alpha_v = 0 \tag{7.16}$$

$$\delta_u - \gamma\alpha_u = 0. \tag{7.17}$$

Now, according to (3.18),

$$e^{2\phi} = \tfrac{1}{4}(e^{-\alpha}\gamma - e^{-\alpha}\delta)^2$$

so that, from (2.13),

$$dS^2 = \tfrac{1}{4}(\gamma - \delta)^2(du^2 + dv^2). \tag{7.18}$$

A comparison of (7.12) and (7.18) shows that

$$\gamma = \delta + 2Q. \tag{7.19}$$

Then integration of (7.17) leads to

$$\delta + 2Q = e^{\alpha}f(v) \tag{7.20}$$

where $f(v)$ is an arbitrary function of v. Substitution of (7.19) and (7.20) in (7.16) and integration results in the equation[3]

---

[3] Since the geodesic curvatures of the lines of curvature of a V_2 satisfying (7.21), v = constant and u = constant, are $-h'(v)$ and $-g'(u)$, respectively, each line of curvature has constant geodesic curvature.

(7.21)
$$e^{-\alpha} = g(u) + h(v)$$
where
(7.22)
$$h'(v) = f'(v)/2Q.$$

If (7.19), (7.20) and (7.21) are substituted in (7.15), we obtain

(7.23)
$$[f^2 - 2Qfh + h'^2 - hh''] \\ + [g'^2 - gg''] - g[2fQ + h''] - hg'' = 0.$$

If (7.23) is differentiated successively with respect to u and v, we obtain

(7.24)
$$g'[(2fQ)' + h'''] + h'g''' = 0.$$

(I) Suppose that $g'h' \neq 0$. Then division of (7.24) by $g'h'$ leads to two separate equations which when integrated yield the equations

(7.25)
$$g'' = c_1 g + c_2$$
(7.26)
$$2fQ + h'' = - c_1 h + c_3$$

where c_1, c_2, c_3 are constants. We distinguish several cases:

(I_1) $c_1 = 0$. The solution of (7.25) is

(7.27)
$$g = \tfrac{1}{2} c_2 u^2 + c_4$$

after a translation of u, if necessary. Substitution of this result and (7.26) in (7.23) leads to

(7.28)
$$c_1 = c_3$$
(7.29)
$$f^2 - 2fhQ + h'^2 - hh'' - c_2 h - 2c_2 c_4 = 0.$$

Use of (7.26) permits us to write (7.29) as

(7.30)
$$f^2 + h'^2 - 2c_2 h - 2c_2 c_4 = 0.$$

According to (7.22), if $f' = 0$ then $h' = 0$. Since this case has been excluded, f' is not zero. Then (7.26) is a consequence of (7.22) and (7.30) which are the final conditions on $f(v)$ and $h(v)$ in this case.

(I_2) $c_1 = - b^2$. The solution of (7.25) is

(7.31)
$$g = c_4 \cos bu + c_2/b^2$$

after a translation of u, if necessary. Substitution of this result and (7.26) in (7.23) yields (7.28) and

(7.32)
$$f^2 - 2fhQ + h'^2 - hh'' + c_4{}^2 b^2 - c_2 h - c_2{}^2/b^2 = 0.$$

By means of (7.26), (7.32) becomes

(7.33)
$$f^2 + h'^2 - b^2 h^2 + c_4{}^2 b^2 - 2c_2 h - c_2{}^2/b^2 = 0.$$

Since f' differs from zero, (7.26) is a consequence of (7.22) and (7.33). These are the final two conditions on $f(v)$ and $h(v)$ in this case.

(I$_3$) $c_1 = + b^2$. The solution of (7.25) is

$$\begin{aligned} &\text{(i)} \quad g = c_4 \cosh bu - c_2/b^2 \\ (7.34) \qquad &\text{(ii)} \quad g = c_4 \sinh bu - c_2/b^2 \\ &\text{(iii)} \quad g = c_4 e^{bu} - c_2/b^2 \end{aligned}$$

after a translation of u, if necessary. If (7.34) is substituted in (7.23) and use is made of (7.26), we find that (7.28) is true and also obtain

$$(7.35) \qquad f^2 - 2fhQ + h'^2 - hh'' - \epsilon c_4{}^2 b^2 - c_2 h + c_2{}^2/b^2 = 0$$

where $\epsilon = +1, -1, 0$ in cases (i), (ii) and (iii), respectively. Using (7.26), (7.35) becomes

$$(7.36) \qquad f^2 + h'^2 + b^2h^2 - \epsilon c_4{}^2 b^2 - 2c_2 h + c_2{}^2/b^2 = 0.$$

As in the previous two cases, (7.22) and (7.36) are the final conditions on $f(u)$ and $h(v)$.

(II) We now assume that $g'h' = 0$. Then (7.24) is true. We consider several cases:

(II$_1$) $g' = h' = 0$. We may assume, without loss of generality, that

$$(7.37) \qquad g = 1, \qquad h = 0.$$

Substitution of (7.37) in (7.23) leads to

$$(7.38) \qquad f = 0$$

which is also consistent with (7.22). Hence (7.37) and (7.38) are the final conditions in this case.

(II$_2$) $g' = 0$, $h' \neq 0$. Hence

$$(7.39) \qquad g = c_4.$$

Then (7.23) becomes, after use of (7.22),

$$(7.40) \qquad f^2 + h'^2 = b^2(c_4 + h)^2.$$

Equations (7.22) and (7.40) are the final conditions.

(II$_3$) $h' = 0$, $g' \neq 0$. Taking account of (7.22),

$$(7.41) \qquad h = c_4, \qquad f = c_5,$$

so that (7.23) becomes

$$(7.42) \qquad c_5{}^2 - 2c_4c_5Q + g'^2 - gg'' - 2c_5gQ - c_4g'' = 0.$$

Successive differentiation with respect to u and v shows that $c_5 = 0$ since $Q' \neq 0$. Hence the solution of (7.42) is

$$(7.45) \qquad g = c_6e^{bu} - c_4.$$

The solutions obtained under the above six categories exhaust the possibilities for a given function $Q(v)$. In all cases, the first and second conformal fundamental forms are the same but the third conformal fundamental form (7.13) may differ in the additive constant c. We shall calculate E_{11} for each of the above solutions in order to distinguish the various conformally different types. From (7.19), (7.20) and (7.21), the principal normal curvatures of any of these surfaces are

$$(7.46) \qquad {}_{(1)}b = f(v), \qquad {}_{(2)}b = f(v) - 2Q[g(u) + h(v)].$$

According to (3.18),

$$(7.47) \qquad e^{\phi} = Q(g+h), \qquad \mu = f - Q(g+h).$$

Also, from (7.12),

$$\Gamma_{11}{}^1 = 0 \qquad \Gamma_{11}{}^2 = -Q'/Q.$$

From these equations, we find that

$$\phi_{:1} = g'/(g+h) \qquad \phi_{:2} = Q'/Q + h'/(g+h)$$
$$\phi_{:11} = [g''(g+h) - g'^2]/(g+h)^2 + Q'^2/Q^2 + 2h'Q'/(g+h).$$

From the second set of equations (2.14) and the above equations,

$$(7.48) \qquad E_{11} = -Q^2/2 - \frac{1}{2}\,\frac{Q'^2}{Q^2} + \tfrac{1}{2}Q^2(e^{-\phi}\mu + 1)^2$$
$$+ [1/(g+h)^2][\tfrac{1}{2}g'^2 - g''(g+h) + \tfrac{1}{2}h'^2].$$

But, from (7.47),

$$e^{-\phi}\mu + 1 = f/Q(g+h).$$

Substitution of this result in (7.48) gives

$$(7.49) \qquad E_{11} = -Q^2/2 - \frac{1}{2}\,\frac{Q'^2}{Q^2} + 1/2(g+h)^2$$
$$\times [g'^2 - 2gg'' - 2hg'' + h'^2 + f^2].$$

Comparison of (7.13) and (7.49) shows that

$$c = [1/2(g+h)^2][g'^2 - 2gg'' - 2hg'' + h'^2 + f^2].$$

If the right member of this equation be evaluated by means of the conditions on $f(v), h(v)$ in each case, we find that

$$(7.50) \qquad \begin{array}{rll} \text{(i)} & c = 0 & \text{in cases } \mathrm{I}_1, \mathrm{II}_1 \\ \text{(ii)} & c = \tfrac{1}{2}b^2 & \text{in cases } \mathrm{I}_2, \mathrm{II}_2 \\ \text{(iii)} & c = -\tfrac{1}{2}b^2 & \text{in cases } \mathrm{I}_3, \mathrm{II}_3. \end{array}$$

It follows from the equivalence theorem of **2** that all surfaces which come under cases I_1 and II_1, or cases I_2 and II_2 for the same b, or cases I_3 and II_3 for the same b are conformally equivalent in R_3. It is only necessary to identify one of the surfaces in each category in order to obtain a complete classification of all isothermic gradient surfaces. This is done below:

(i) $c = 0$. For a surface belonging to II_1, (7.14) becomes, after using (7.37), (7.38), (7.19), (7.20) and (7.21),

$$ds^2 = du^2 + dv^2$$
$$b_{ij}dx^i dx^j = -2Q(v)\,dv^2.$$

It is clear from these forms that the surface V_2 is developable, that v denotes ordinary distance along one family of lines of curvature and that $-2Q(v)$ is the principal normal curvature of V_2 along each line of the family. Hence V_2 is a right cylindrical surface whose directrix is the plane curve defined intrinsically by

$$(7.51) \qquad k = -2(s).$$

Here k and s are the curvature and arc length of the directrix.

(ii) $c = \frac{1}{2}b^2$. A surface of type II_2 has the fundamental forms

$$(7.52) \qquad \begin{aligned} ds^2 &= (1/h^2)(du^2 + dv^2) \\ b_{ij}dx^i dx^j &= (f/h^2)\,du^2 + (f/h^2 - 2Q/h)\,dv^2 \end{aligned}$$

where $f(v), h(v)$ obey (7.22) and (7.40). Since the coefficients in (7.52) are all functions of v only, comparison with a surface of revolution is suggested. Now the equations of any surface of revolution are

$$y^1 = \omega \cos \alpha, \qquad y^2 = \omega \sin \alpha, \qquad y^3 = \nu(\omega).$$

From these equations, we obtain the following fundamental forms

$$(7.53) \qquad \begin{aligned} ds^2 &= \omega^2 d\alpha^2 + \sec^2 \theta \cdot d\omega^2 \\ b_{ij}dx^i dx^j &= \omega \sin \theta \cdot d\alpha^2 + \sec \theta \cdot (d\theta/d\omega) \cdot d\omega^2 \end{aligned}$$

where $\theta(\omega) = \arctan (d\nu/d\omega)$. We introduce the isothermal parameter χ by the equation

$$\chi = \int (\sec\theta/\omega)\, d\omega. \tag{7.54}$$

If (7.52) and (7.53) refer to the same surface

$$u = d\alpha, \qquad v = d\chi \tag{7.55}$$

where d is a constant. Equating the coefficients of the respective fundamental forms of (7.52) and (7.53), we obtain the equations

$$h = d/\omega, \qquad f = \sin\theta/\omega \tag{7.56}$$

$$\omega^2 \cos\theta (d\theta/d\omega) = d^2(f/h^2 - 2Q/h). \tag{7.57}$$

From (7.54), (7.55) and (7.56), we obtain

$$\begin{aligned} dh/dv &= \frac{1}{c}\,\frac{dh}{d\chi} = (1/c)\,\omega\cos\theta(dh/d\omega) = -\cos\theta/\omega \\ df/dv &= (1/c)\omega\cos\theta(df/d\omega) \\ &= (1/c)\left[\cos^2\theta(d\theta/d\omega) - (1/\omega)\cos\theta\sin\theta\right]. \end{aligned} \tag{7.58}$$

Now h and f satisfy (7.22) and (7.40). Substitution of (7.56) and (7.58) in (7.40) yields $d = 1/b$. Similarly (7.22) reduces to (7.57) which may be simplified by means of (7.56) to give

$$2dQ(d\chi) = \sin\theta - \omega\cos\theta(d\theta/d\omega). \tag{7.59}$$

If $Q(v)$ is given, equations (7.54) and (7.59) may always be solved for $\theta(\omega)$. Conversely, if $\theta(\omega)$ is given, (7.54) and (7.59) define $Q(v)$. Hence the V_2 defined by (7.52) is a surface of revolution and every surface of revolution is of type II_2.

(iii) $c = -\frac{1}{2}b^2$. A surface of type II_3 has the fundamental forms

$$\begin{aligned} ds^2 &= e^{2bu}(du^2 + dv^2) \\ b_{ij}dx^i dx^j &= -2Qe^{bu}dv^2. \end{aligned} \tag{7.60}$$

By means of the parameter transformation

$$u' = (1/b)e^{bu}, \qquad v' = bv \tag{7.61}$$

(7.60) becomes

$$\begin{aligned} ds^2 &= du'^2 + u'^2 dv'^2 \\ b_{ij}dx^i dx^j &= -(2/b)u'Q(v'/b)\cdot dv'^2. \end{aligned} \tag{7.62}$$

K

The first form of (7.62) is identical with that of the plane if the parameters are polar coördinates. Hence the surface V_2 defined by (7.62) is a developable surface whose lines of curvature are applicable to a plane system of polar coördinates. Here u', v' correspond to the length of the radius vector and the vectorial angle respectively. This means that the V_2 defined by (7.62) must be a conical surface. The converse is also true. The line of curvature, $u' = 1/b$, may be considered as the intersection of the conical surface V_2 and the sphere whose center is at the point $u' = 0$ and whose radius is $1/b$. Classical results of differential geometry lead to the conclusion that the geodesic curvature of the intersection curve relative to the sphere is the normal curvature in its direction relative to the conical V_2. But, using (7.62), the normal curvature of $u' = 1/b$ is given by

$$ {}_{(2)}b = -2Q(v)/bu' = -2Q(v). $$

Also, from (7.61), since dv' is a small central angle, dv is an element of distance on the sphere. Hence V_2 is the conical surface whose vertex is at the center of the sphere of radius $1/b$ and whose directrix on this sphere is defined by the intrinsic equation

$$ k = -2Q(s). $$

Here k and s are the geodesic curvature and arc length of the spherical curve. We summarize these results in the following theorem:

THEOREM 7.2. *Every isothermic gradient surface is conformally equivalent to a conical surface, cylindrical surface or a surface of revolution, and conversely.*

The generators of a cylindrical surface may be considered as circles which are all tangent to each other at the point at infinity. Similarly, the generators of a conical surface may be considered as circles which all have two common points—the vertex and the point at infinity. Consequently an isothermic gradient surface is of type (i), (ii), (iii) of (7.50) according as the members of the family of circular lines of curvature have 1, 0 or 2 common points. In particular, this topological classification applies to Dupin cyclides.

8. Gradient surfaces of constant conformal Gaussian curvature. We consider a gradient surface V_2 for which J is constant. Let the lines of curvature be parametric. Since V_2 is gradient, in accordance with the results of **4**, we may choose the components of its conformal measure tensor as

$$ X = e^{2Q}, \quad Y = e^{-2Q}. \tag{8.1} $$

The condition (3.12) is satisfied in virtue of (8.1). Using (4.1) and (4.8), the remaining conditions (3.11), (3.13) and (3.14) may be written as

$$ {}_{(1)}E + {}_{(2)}E = J + 1. \tag{8.2} $$

(8.3) $${}_{(1)}E_v = -Q_v(E' + 2)$$

(8.4) $${}_{(2)}E_u = -Q_u(E' - 2).$$

From (4.8) and (8.2),

(8.5) $${}_{(1)}E = \tfrac{1}{2}[1 + J + E'], \qquad {}_{(2)}E = \tfrac{1}{2}[1 + J - E'].$$

Since J is constant, if these values are substituted in (8.3) and (8.4) we obtain

(8.6) $$E'_v = -2Q_v(E' + 2)$$

(8.7) $$E'_u = 2Q_u(E' - 2).$$

We consider the possible solutions of (8.6) and (8.7).

(I) Suppose that Q is a constant. Then $J = 0$ and V_2 is a Dupin cyclide.

(II) Suppose that Q is a non-constant function[4] $Q(v)$ of v only. Then, from (8.7), $E' = E'(v)$. Integration of (8.6) yields

(8.8) $$E' = c_1 e^{-2Q} - 2$$

where c_1 is a constant. The remaining condition, $J =$ constant, leads to

(8.9) $$e^{2Q} = -Jv^2 + c_2 v + c_3$$

where the c's are constants, after use of (3.7) and (3.10). We consider three cases:

(II$_1$) $J = 0$. We may suppose that $c_2 \neq 0$, the other alternative having been considered in (I). Then the substitution of (8.9) in (8.1) followed by the parameter transformation

$$u' = c_2 u/2, \qquad v' = (2/c_2)(c_2 v + c_3)^{1/2}$$

permits us to write the first conformal fundamental form of V_2 as

(8.10) $$dS^2 = v^2 du^2 + dv^2.$$

The primes have been omitted in (8.10). The deviation vector and the remaining conformal fundamental forms are readily found by reference to (3.2), (3.5), (8.2) and (8.8). They are

$$F_1 = 0 \qquad F_2 = -2/v$$

(8.11) $$B_{ij}dx^i dx^j = v^2 du^2 - dv^2$$

$$E_{ij}dx^i dx^j = (-\tfrac{1}{2}v^2 + d)\,du^2 + (3/2 - d/v^2)\,dv^2.$$

---

[4] The case $Q = Q(u)$ is equivalent to the one discussed here after a parameter transformation which interchanges the roles of u and v.

where d is an arbitrary constant.

(II$_2$) $J = a^2$. In this case, we may write (8.9) as

$$e^{2Q} = c_4{}^2 - (av + c_5)^2. \tag{8.12}$$

If we substitute (8.12) in (8.1) and employ the parameter substitution

$$u' = c_4 u, \qquad v' = (1/a) \arcsin [(av + c_5)/c_4]$$

the first conformal fundamental form of V_2 becomes[5]

$$dS^2 = \cos^2 av du^2 + dv^2 \tag{8.13}$$

after dropping the primes. The other forms and F_i are found as in (II$_1$) and are

$$\begin{aligned} B_{ij}dx^i dx^j &= \cos^2 av du^2 - dv^2 \\ E_{ij}dx^i dx^j &= [\tfrac{1}{2}(a^2 - 1) \cos^2 av + d]du^2 \\ &\quad + [\tfrac{1}{2}(a^2 + 3) - d \sec^2 av]dv^2 \\ F_1 = 0, &\qquad F_2 = 2a \tan av. \end{aligned} \tag{8.14}$$

(II$_3$) $J = -a^2$. We may now write (8.9) as

$$e^{2Q} = (av + c_5)^2 \pm c_4{}^2. \tag{8.15}$$

We substitute (8.15) in (8.1) and then use the parameter transformation

$$\begin{array}{lll} \text{(i)} & u' = c_4 u, & v' = (1/a) \operatorname{arc\,cosh} [(av + c_5)/c_4] \\ \text{(ii)} & u' = c_4 u, & v' = (1/a) \operatorname{arc\,sinh} [(av + c_5)/c_4] \\ \text{(iii)} & u' = u, & v' = (1/a) \log (av + c_5) \end{array}$$

according as (8.15) contains the plus sign or minus sign or $c_4 = 0$. In this manner, we obtain three possible forms:

$$\begin{array}{ll} \text{(i)} & dS^2 = \cosh^2 av du^2 + dv^2 \\ \text{(ii)} & dS^2 = \sinh^2 av du^2 + dv^2 \\ \text{(iii)} & dS^2 = e^{2av} du^2 + dv^2 \end{array} \tag{8.16}$$

after dropping the primes. The other forms and F_i are

$$\begin{aligned} B_{ij}dx^i dx^j &= X du^2 - dv^2 \\ E_{ij}dx^i dx^j &= [-\tfrac{1}{2}(1 + a^2)X + d]du^2 + [\tfrac{1}{2}(3 - a^2) - dX^{-1}]dv^2 \\ F_1 = 0, &\qquad F_2 = -(\log X)_v \end{aligned} \tag{8.17}$$

where X equals $\cosh^2 av$, $\sinh^2 av$, e^{2av} in cases (i), (ii), (iii), respectively.

---

[5] Equation (8.13) may also be written

$$dS^2 = \sin^2 av \cdot du^2 + dv^2.$$

A comparison of (7.10) with (8.11), (8.14) and (8.17) respectively reveals that

$$d = \begin{cases} \tfrac{1}{2}(2c - 1), & J = 0 \\ \tfrac{1}{2}(2c - a^2), & J = a^2 \\ \tfrac{1}{2}(2c + a^2), & J = -a^2, \text{ case (i)} \\ \tfrac{1}{2}(2c - a^2), & J = -a^2, \text{ case (ii)} \\ c, & J = -a^2, \text{ case (iii).} \end{cases} \tag{8.18}$$

(III) Next let us suppose that both variables u, v are effectively present in Q so that $Q_u \cdot Q_v \neq 0$. It follows from (8.6) and (8.7) that E' cannot be 2 or -2 and indeed must be a function of both u and v. Integration of (8.6) and (8.7) leads to

$$E' + 2 = g_1(u) \cdot e^{-2Q}, \qquad E' - 2 = h_1(v) \cdot e^{2Q}.$$

Since we obtain

$$E' = [g_1 h_1 + 4]^{1/2}$$

from these equations it follows that neither g_1 nor h_1 is constant. We also find that

$$e^{2Q} = (E' - 2)/h_1 = g_1/(E' + 2).$$

Hence the first conformal fundamental form of V_2 may be written as

$$dS^2 = \frac{[g_1 h_1 + 4]^{1/2} - 2}{h_1}\, du^2 + \frac{[g_1 h_1 + 4]^{1/2} - 2}{g_1}\, dv^2. \tag{8.19}$$

After the parameter transformation

$$u' = \tfrac{1}{2} g_1(u), \qquad v' = \tfrac{1}{2} h_1(v)$$

(8.19) becomes

$$dS^2 = \frac{[uv + 1]^{1/2} - 1}{v g_2(u)}\, du^2 + \frac{[uv + 1]^{1/2} + 1}{u h_2(v)}\, dv^2 \tag{8.20}$$

where we have written

$$g_2(u') = \tfrac{1}{4} g_1'^2[u(u')], \qquad h_2(v') = \tfrac{1}{4} h_1'^2[v(v')]$$

and dropped the primes.

It remains to impose the condition that the V_2 having (8.20) as its first conformal fundamental form has a constant value for J (or K). We shall prove that it is impossible to satisfy this condition. From (3.7) and (8.20),

$$\begin{aligned} K = {} & \frac{(3u^2v^2 + 12uv + 8)}{4(uv + 1)^{3/2}} \cdot (g_2/u^3 + h_2/v^3) - \frac{(uv + 2)}{4(uv + 1)^{1/2}} \cdot (g'_2/u^2 + \\ & + 2g_2/u^3 - 2h_2/v^3 - g'_2/2u^2 + h'_2/2v^2. \end{aligned} \tag{8.21}$$

We define $g(u)$, $h(v)$ by

$$g = g_2/u^3, \qquad h = h_2/v^3.$$

Then

$$g'_2/u^2 = 3g + ug', \qquad h'_2/v^2 = 3h + vh'.$$

Hence (8.21) may be written as

$$K = \frac{3uv + 2}{4(uv + 1)^{3/2}}(g + h) - \frac{uv + 2}{4(uv + 1)^{1/2}}(ug' + vh') \\ - g + h - ug' + vh'. \tag{8.22}$$

The first two expressions in the right member of this equation are of the form

$$e(u, v) = f(z)[u^n g^{(n)} + v^n h^{(n)}]$$

where (n) denotes n-fold differentiation and $z = uv$. Successive differentiation of this equation gives

$$e_{uv} = (zf'' + (1 + n)f')[u^n g^{(n)} + v^n h^{(n)}] \\ + f'[u^{n+1} g^{(n+1)} + v^{n+1} h^{(n+1)}]. \tag{8.23}$$

where the prime indicates differentiation with respect to z. We differentiate (8.22) successively with respect to u and v and apply (8.23). After multiplication by $16(z + 1)^{7/2}$, we obtain

$$(3z^2 - 12z)(g + h) - (3z^3 + 15z^2 + 12z)[ug^{(1)} + vh^{(1)}] \\ - (2z^3 + 4z^2 + 2z)[u^2 g^{(2)} + v^2 h^{(2)}] = 0.$$

If we apply (8.23) to this equation three more times we are led to the linear differential equation

$$72[ug^{(1)} + vh^{(1)}] + 228[u^2 g^{(2)} + v^2 h^{(2)}] + 156[u^3 g^{(3)} + v^3 h^{(3)}] \\ + 33[u^4 g^{(4)} + v^4 h^{(4)}] + 2[u^5 g^{(5)} + v^5 h^{(5)}] = 0.$$

From this equation, we conclude that

$$72ug^{(1)} + 228u^2 g^{(2)} + 156u^3 g^{(3)} + 33u^4 g^{(4)} + 2u^5 g^{(5)} = c_1$$
$$72vh^{(1)} + 228v^2 h^{(2)} + 156v^3 h^{(3)} + 33v^4 h^{(4)} + 2v^5 h^{(5)} = -c_1.$$

Integration of these differential equations by the customary methods leads to

$$g(u) = (c_1/6)\log u + c_2 + c_3 u^{-1/2} + c_4 u^{-1} + c_5 u^{-2} + c_6 u^{-3}$$

and a similar solution for $h(v)$. But substitution of these solutions in (8.22) readily leads to a contradiction. Hence functions $g(u)$, $h(v)$ which satisfy

(8.22) do not exist. This means that there are no surfaces which obey the requirements of case (III). Since the first conformal fundamental form for each surface which occurs under case (I) or case (II) satisfies (7.9), it follows from Theorem 7.1 that these surfaces are all isothermic gradient. We state these results in the following theorem:

THEOREM 8.1. *If a gradient surface has constant conformal Gaussian curvature J, then it is isothermic. If $J = 0$, the surface is a Dupin cyclide or its conformal fundamental forms are defined by* (8.10) *and* (8.11) *in some parameter system. If J equals $+a^2$ or $-a^2$ its conformal fundamental forms are defined by* (8.13) *and* (8.14) *or by* (8.16) *and* (8.17), *respectively, in some parameter system.*

As noted in **6**, it follows from (3.11) and (4.1) that a surface V_2 whose principal deviations are each constant has constant conformal Gaussian curvature. Consequently if V_2 is also a gradient surface, its conformal fundamental forms are defined by Theorem 8.1. In this case, unless V_2 is a Dupin cyclide, the only constant values of ${}_{(1)}E$ and ${}_{(2)}E$ are obtained by equating d to zero in (8.11), (8.14) and (8.17). Hence *a gradient surface having constant principal deviations is either a Dupin cyclide or its principal deviations are*

$$(8.24) \qquad {}_{(1)}E = -1/2 + J/2, \qquad {}_{(2)}E = 3/2 + J/2.$$

According to the results of **7**, each of these surfaces is conformally equivalent to a cylindrical surface, surface of revolution or conical surface according as c equals, is greater than or is less than zero. Since $d = 0$, it follows from (8.18) that a surface (which is not a Dupin cyclide) having constant principal deviations is conformally equivalent to a surface of revolution if J is zero or positive. If J is negative, the surface is conformal to a conical surface, surface of revolution or cylindrical surface in cases (i), (ii) or (iii) of (8.16) respectively.

As an example of such a surface, we mention the catenoid referred to in **6** for which $J = 1$. We also note the right cylindrical surface whose directrix is the equiangular spiral $\rho = e^{(a/2)\theta}$. The intrinsic equation of the spiral is $k = 2/as$. Comparison with (7.51) shows that the first conformal fundamental form of the cylindrical surface is

$$dS^2 = (1/a^2v^2)(du^2 + dv^2)$$

which is equivalent to (iii) of (8.16) after the introduction of isothermal parameters in the latter.

9. Isothermic surfaces of constant conformal Gaussian curvature. If

V_2 is isothermic, the first and second conformal fundamental forms may be written as

$$\begin{aligned} dS^2 &= e^{2Q}(du^2 + dv^2) \\ B_{ij}dx^i dx^j &= e^{2Q}(du^2 - dv^2). \end{aligned} \tag{9.1}$$

In this parameter system, from (3.7) and (3.10),

$$J = -e^{-2Q}(Q_{uu} + Q_{vv}). \tag{9.2}$$

The conditions (3.11) to (3.14) become

$$E_{11} + E_{22} = e^{2Q}(J+1), \tag{9.3}$$

$$E_{12} = -2Q_{uv} \tag{9.4}$$

$$\partial E_{11}/\partial v = -2Q_{uuv} + e^{2Q}Q_v(J-1) \tag{9.5}$$

$$\partial E_{22}/\partial u = -2Q_{uvv} + e^{2Q}Q_u(J-1) \tag{9.6}$$

respectively. We also assume that J is constant. Then, as a consequence of (9.3) and (9.6), we obtain

$$\partial E_{11}/\partial u = 2Q_{uvv} + e^{2Q}Q_u(J+3). \tag{9.7}$$

The integrability condition of (9.5) and (9.7) may be written as

$$[Q_{uu} + Q_{vv}]_{uv} + 2e^{2Q}(2Q_uQ_v + Q_{uv}) = 0$$

or, using (9.2),

$$e^{2Q}(1-J)(2Q_uQ_v + Q_{uv}) = 0. \tag{9.8}$$

We first suppose that $J \neq 1$. Then (9.8) yields

$$e^{2Q} = Q_1(u) + Q_2(v). \tag{9.9}$$

By virtue of (9.9), (9.5) and (9.6) may be integrated. This gives

$$\begin{aligned} E_{11} &= -2Q_{uu} + \tfrac{1}{2}(J-1)e^{2Q} + g(u), \\ E_{22} &= -2Q_{vv} + \tfrac{1}{2}(J-1)e^{2Q} + h(v). \end{aligned} \tag{9.10}$$

Substitution of these results in (9.2) and use of (9.9) leads to

$$g(u) = c - 2(J-1)Q_1, \qquad h(v) = -c - 2(J-1)Q_2.$$

If these values are substituted in (9.10), we obtain

$$\begin{aligned} E_{11} &= -2Q_{uu} + \tfrac{1}{2}(J-1)(Q_2 - 3Q_1) + c \\ E_{22} &= -2Q_{vv} + \tfrac{1}{2}(J-1)(Q_1 - 3Q_2) - c. \end{aligned} \tag{9.11}$$

It only remains to satisfy (9.2). In this way, from (9.2) and (9.9),

$$Q_1''Q_1 - Q_1'^2 + Q_2''Q_2 - Q_2'^2 + Q_2Q_1'' + Q_1Q_2'' = -2J(Q_1 + Q_2)^3 \tag{9.12}$$

where the primes indicate differentiation, with respect to the independent variable, of each function. If $Q'_1Q'_2 = 0$, we may suppose, after a parameter change, that $Q_1 = 0$. Then (9.12) becomes

$$Q_2''Q_2 - Q'^2_2 + 2JQ_2^3 = 0. \tag{9.13}$$

In accordance with Theorem 7.1, the surfaces for which $Q_1 = 0$ must be isothermic gradient surfaces with constant J. They are discussed in **8**. The first conformal fundamental forms of these surfaces, when the parameters are isothermal, may be derived from (9.13) or, more simply, by restating (6.1), (8.10), (8.13) and (8.16) in an isothermal parameter system. If $Q'_1Q'_2 \neq 0$, successive differentiation of (9.12) with respect to u and v and division by $Q'_1Q'_2$ gives

$$Q_1'''/Q'_1 + 12JQ_1 + Q_2'''/Q_2' + 12JQ_2 = 0$$

which, after separation and integration, leads to

$$\begin{aligned} Q_1'' &= -6JQ_1^2 + cQ_1 + c_1 \\ Q''_2 &= -6JQ_2^2 - cQ_2 + c_2 \end{aligned} \tag{9.14}$$

where the c's are constants.

We consider various cases:

(I) $J = 0$. The solutions of (9.13) define first conformal fundamental forms which are equivalent to (6.1) and (8.10). These forms are

$$\begin{aligned} dS^2 &= du^2 + dv^2 \\ dS^2 &= e^{2v}(du^2 + dv^2) \end{aligned} \tag{9.15}$$

respectively.

We now consider the solution of (9.14) and assume that $c = 0$. Then

$$Q_1 = (c_1/2)u^2 + c_3u + c_4, \qquad Q_2 = (c_2/2)v^2 + c_5v + c_6.$$

Substitution of these results in (9.12) gives the additional conditions

$$c_1 = c_2, \qquad 2c_1(c_4 + c_6) = c_3^2 + c_5^2.$$

Hence

$$e^{2Q} = (c_1/2)[(u + c_7)^2 + (v + c_8)^2].$$

After the parameter transformation

$$u' = (c_1/2)^{1/4}(u + c_7), \qquad v' = (c_1/2)^{1/4}(v + c_8)$$

and dropping primes, (9.1) becomes

$$dS^2 = (u^2 + v^2)(du^2 + dv^2). \tag{9.16}$$

Now suppose that $c \neq 0$ in (9.14). If $c = b^2$, integration of (9.14) gives

$$Q_1 = \begin{cases} c_3 \cosh (bu + c_4) - c_1/b^2 \\ c_3 \sinh (bu + c_4) - c_1/b^2 \\ c_3 e^{bu} - c_1/b^2 \end{cases}$$

$$Q_2 = \begin{cases} c_5 \cos (bv + c_6) + c_2/b^2 \\ c_5 \sin (bv + c_6) + c_2/b^2. \end{cases}$$

The solutions for Q_2 are equivalent to each other while those for Q_1 are not. If the first (or second) solution of each set is substituted in (9.12), we obtain

$$c_1 = c_2, \qquad c_3 = \pm c_5.$$

After the parameter transformation

$$u' = bu + c_4, \qquad v' = bv + c_6$$

and writing $c_3/b^2 = a^2$ and dropping primes, the forms (9.1) corresponding to these solutions are

$$\begin{aligned} dS^2 &= a^2 (\cosh u \pm \cos v)(du^2 + dv^2) \\ dS^2 &= a^2 (\sinh u \pm \sin v)(du^2 + dv^2). \end{aligned} \tag{9.17}$$

If the third solution for Q_1 and either solution for Q_2 is substituted in (9.12), we find that $c_5 = 0$. Hence this solution is excluded since Q_2 must actually involve v. Finally, we note that if $c = -b^2$ in (9.14), u and v interchange rôles so that the resulting solution is equivalent to (9.17) after a change of parameter.

(II) $J = a^2$. The first conformal fundamental form defined by the solutions of (9.13) is equivalent to (8.13). In this way, we obtain

$$dS^2 = \operatorname{sech}^2 av (du^2 + dv^2). \tag{9.18}$$

We now consider the solution of (9.14). Integration leads to

$$\begin{aligned} Q'^2_1 &= -4a^2 Q_1^3 + cQ_1^2 + 2c_1 Q_1 + c_3 \\ Q'^2_2 &= -4a^2 Q_2^3 - cQ_2^2 + 2c_2 Q_2 + c_4 \end{aligned} \tag{9.19}$$

where the c's are constants. If these equations and (9.14) are substituted in (9.12), we find that

$$c_1 = c_2, \qquad c_3 = -c_4$$

so that, after the substitution

$$Q_1 = Q_3 + c/12a^2, \qquad Q_2 = Q_4 - c/12a^2$$

(9.19) may be written as

$$Q'^2_3 = -a^2[4Q_3^2 - g_2Q_3 - g_3]$$
$$Q'^2_4 = -a^2[4Q_4^2 - g_2Q_4 + g_3]$$

where the g's are constants. The solutions of these equations, after translation of u and v, if necessary, are

$$(9.20)\qquad \begin{aligned} Q_1 &= \wp(iau \mid g_2, g_3) + c/12a^2 \\ Q_2 &= \wp(iav \mid g_2, -g_3) - c/12a^2 \end{aligned}$$

where $\wp(x \mid g_2, g_3)$ is the Weierstrass $\wp$-function of x with the invariants g_2, g_3. Since $\wp(x)$ is an even function and g_2, g_3 are real, the functions Q_1 and Q_2 defined by (9.20) are real for real values of u and v. The first conformal fundamental form corresponding to (9.20) is

$$(9.21)\qquad dS^2 = [\wp(iau \mid g_2, g_3) + \wp(iav \mid g_2, g_3)](du^2 + dv^2).$$

(III) $J = -a^2$. The solutions of (9.13) define first conformal fundamental forms which are equivalent to (8.16). These forms are

$$(9.22)\qquad \begin{aligned} &\text{(i)} && dS^2 = \sec^2 av(du^2 + dv^2) \\ &\text{(ii)} && dS^2 = \operatorname{csch}^2 av(du^2 + dv^2) \\ &\text{(iii)} && dS^2 = (1/a^2v^2)(du^2 + dv^2). \end{aligned}$$

If we proceed wth (9.14) as in (II) above, we obtain

$$(9.23)\qquad dS^2 = [\wp(au \mid g_2, g_3) + \wp(av \mid g_2, -g_3)](du^2 + dv^2)$$

as the first conformal fundamental form corresponding to the solution.

(IV) We now assume that $J = 1$. Then (9.8) is true. In this case, any solution $Q(u, v)$ of (9.2) with $J = 1$ defines the first conformal fundamental form of an actual surface. The components of E_{ij} are then defined by (9.4) and by

$$E_{11} = -2Q_{uu} + c, \qquad E_{22} = -2Q_{vv} - c.$$

These last equations are special cases of (9.11).

We summarize these results as follows:

THEOREM 9.1. *An isothermic surface of constant conformal curvature J ($J \neq 1$) is either a gradient surface*[6] *or its first conformal fundamental form is defined by* (9.16) *or* (9.17) *if $J = 0$, by* (9.21) *if $J = a^2$ and by* (9.23) *if $J = -a^2$. If $J = 1$, any solution of* (9.2) *defines the first con-*

---

[6] In this case, its conformal fundamental forms are given by Theorem 8.1.

formal fundamental form of a surface. The components of the other conformal fundamental forms in all cases are defined by (9.1), (9.4) *and* (9.11).

10. Surfaces with indeterminate lines of deviation. If a surface V_2 is homogeneous with respect to the deviation tensor,

$$E_{ij} = (E/2)\, G_{ij}. \tag{10.1}$$

In this case, every direction is a deviation direction. We say V_2 has *indeterminate lines of deviation.* It follows from Theorem 4.2 that V_2 is a gradient surface. Hence the components of its conformal measure tensor may be chosen as (8.1) and the remaining conditions are (8.2), (8.3) and (8.4). However, in virtue of (4.1), (4.8) and (10.1),

$$_{(1)}E = {}_{(2)}E = E/2, \qquad E' = 0. \tag{10.2}$$

Therefore (8.3) and (8.4) become

$$E_v = -4Q_v, \qquad E_u = 4Q_u. \tag{10.3}$$

The integrability conditions of these equations are $Q_{uv} = 0$ so that Q may be written as

$$Q = \log Q_2(v) - \log Q_1(u)\,; \qquad e^Q = Q_2/Q_1. \tag{10.4}$$

As a result of (10.3) and (10.4),

$$E = -4 \log Q_1 Q_2 + c \tag{10.5}$$

where c is a constant. If (10.2) and (10.5) are substituted in (8.2), we obtain

$$J = -4 \log Q_1 Q_2 + c - 1.$$

But, from (3.7) and (3.10),

$$J = -\tfrac{1}{2}[Q_1''/Q_2 + Q_2''/Q_1]$$

where the primes indicate differentiation of the functions with respect to their independent variables. From the last two equations,

$$Q_1''/Q_2 + Q_2''/Q_1 = 8 \log Q_1 Q_2 - \tfrac{1}{2}(c - 1). \tag{10.6}$$

Successive differentiation of (10.6) with respect to u and v yields

$$Q_1'''(1/Q_2)' + Q_2'''(1/Q_1)' = 0.$$

If $Q'_1 Q'_2 \neq 0$, this equation leads to two separate equations which integrate as

$$\begin{aligned} Q_1'' &= c_1 \cdot 1/Q_1 + c_2 \\ Q_2'' &= -c_1(1/Q_2) + c_3 \end{aligned} \tag{10.7}$$

where the c's are constants. Substitution of (10.7) in (10.6) leads to

$$c_2/Q_2 + c_3/Q_1 = 8 \log Q_1 + 8 \log Q_2 - \tfrac{1}{2}(c-1)$$

which is impossible.

Therefore $Q'_1 Q'_2 = 0$ and we may suppose, after a parameter change, that $Q_1 = 1$. Hence V_2 is isothermic. The only condition on Q_2 is (10.6) which becomes

$$Q_2'' = 8 \log Q_2 - \tfrac{1}{2}(c-1). \tag{10.8}$$

Any solution of (10.8) together with (10.4) and (8.1) define the components of the first conformal fundamental form of the surface. The only solution of (10.8) for which E is constant is $Q_2 = 1$, $c = 1$ as follows from (10.5). In this case, $E = 1$. The corresponding V_2 is a Dupin cyclide. A comparison of (10.1) with (6.8) shows that V_2 is conformally equivalent to a torus for which $a = \sqrt{2}d$. This torus is characterized by the fact that the tangents drawn from the origin to its generating circle are orthogonal. We state some of these results in the following theorem:

THEOREM 10.1. *Every surface with indeterminate lines of deviation is isothermic. The deviation of the surface is constant if and only if it is conformally equivalent to a torus whose generating circle subtends a right angle at the projection of its center on the axis of rotation.*

11. Gradient hypersurfaces. We call a hypersurface V_n in a Euclidean space R_{n+1} a *gradient hypersurface* if F_i is a gradient on V_n. In this case, there exists a single parameter family of subspaces in V_n, whose equations may be written as $F =$ constant, to which the vector F_i is normal. In this section we derive an analogue of Theorem 4.2 for V_n. The necessary and sufficient condition that F_i be a gradient $F_{:i}$ is that

$$F_{j:k} - F_{k:j} = 0. \tag{11.1}$$

According to (2.22), (11.1) is equivalent to

$$B_{hj}E_k{}^h - B_{hk}E_j{}^h = 0. \tag{11.2}$$

From (2.9) and (2.15), it follows that the n mutually orthogonal principal directions ${}_{(1)}L_i, {}_{(2)}L_i, \cdots, {}_{(n)}L_i$ determined by the tensor B_{ij} relative to G_{ij}

coincide with those determined by b_{ij} relative to g_{ij}. Therefore they are tangent to the lines of curvature of V_n. We also refer to them as the *curvature directions.* The corresponding roots ${}_{(1)}B, {}_{(2)}B, \cdots, {}_{(n)}B$ of the determinantal equation $| B_{ij} - \rho G_{ij} | = 0$ are the *principal conformal normal curvatures* of V_n. The tensor B_{ij} then has the canonical representation

$$B_{ij} = {}_{(1)}B\, {}_{(1)}L_i\, {}_{(1)}L_j + {}_{(2)}B\, {}_{(2)}L_i\, {}_{(2)}L_j + \cdots + {}_{(n)}B\, {}_{(n)}L_i\, {}_{(n)}L_j. \tag{11.3}$$

Now the tensor equations (11.2) are equivalent to the set of scalar equations obtained by multiplying (11.2) by ${}_{(g)}L^i\, {}_{(l)}L^k; g, l = 1, 2, \cdots, n$. Thus (11.2) becomes

$$[{}_{(g)}B - {}_{(l)}B] E_{hj}\, {}_{(g)}L^h\, {}_{(l)}L^j = 0 \tag{11.4}$$

after use of (11.3). There are two ways in which (11.4) may be satisfied. (1) Suppose that ${}_{(g)}B$ is a simple principal conformal normal curvature; that is, ${}_{(g)}B \neq {}_{(l)}B$ for a fixed g and any l not equal to g. Then (11.4) implies that ${}_{(g)}L^h$ is a principal direction of E_{hj} relative to G_{hj}. This means that ${}_{(g)}L^h$ is a deviation direction. (2) Suppose that ${}_{(g)}B$ is a multiple principal conformal normal curvature so that ${}_{(g)}B = {}_{(l)}B$ for $g, l = 1, 2, \cdots, w$. Then the known theory and (11.4) imply that w principal directions of E_{ij} lie in the w-dimensional linear vector space determined by ${}_{(1)}L^i, {}_{(2)}L^i, \cdots, {}_{(w)}L^i$. Due to the symmetry of (11.2), the rôles of E_{ij}, B_{ij} may be interchanged in the above remarks. This completes the proof of the following analogue of Theorem 4.2:

THEOREM 11.1. *If V_n is a gradient hypersurface then* (1) *the curvature direction corresponding to a simple principal conformal normal curvature is also a deviation direction* (2) *the w-dimensional linear vector space of curvature directions corresponding to a principal conformal normal curvature of multiplicity w contains w deviation directions, and conversely.*

The above theorem remains valid if the words "principal conformal normal curvature," "curvature direction" are interchanged with "principal deviation," "deviation direction" respectively.

From (2.19) and (2.20), we obtain

$$K_{ij} = B_{ik} B_{hj} G^{hk} + (2 - n) E_{ij} - E G_{ij} \tag{11.5}$$

after multiplication by G^{hk} and summation on repeated indices. The principal directions determined by the *conformal Ricci* (*surface*) *tensor* K_{ij} are the *conformal Ricci* (*surface*) *directions*. The tensor $B_{ik}B_{hj}G^{hk}$ has the canonical representation

$$B_{ik}B_{hj}G^{hk} = {}_{(1)}B^2\,{}_{(1)}L_i\,{}_{(1)}L_j + {}_{(2)}B^2\,{}_{(2)}L_i\,{}_{(2)}L_j + \cdots + {}_{(n)}B^2\,{}_{(n)}L_i\,{}_{(n)}L_j$$

as follows from (11.3). Hence the principal directions determined by $B_{ik}B_{hj}G^{hk}$ are tangent to the lines of curvature of V_n. As a consequence of this remark and (11.5), we have

THEOREM 11.2. *If a vector is a member of two of the following sets*

(1) *the curvature directions of* V_m

(2) *the deviation directions of* V_n

(3) *the conformal Ricci* (*surface*) *directions of* V_n

and n exceeds 2, it is also a member of the third.

In particular, this theorem has obvious applications to gradient hypersurfaces, as may be seen by reference to Theorem 11.1.

12. Hypersurfaces with indeterminate lines of deviation. A V_n has *indeterminate lines of deviation* if and only if

$$E_{ij} = (E/n)\,G_{ij}. \tag{12.1}$$

If (12.1) is substituted in (2.22), we obtain (11.1). Hence *any* V_n *with indeterminate lines of deviation is a gradient hypersurface.* We may therefore write F_i as $F_{:i}$. Substitution of (12.1) in (2.23) yields

$$E_{:k}\,G_{ij} - E_{:j}\,G_{ik} = n[B_{ij}F_{:k} - B_{ik}F_{:j}]. \tag{12.2}$$

Multiplication of (12.2) by G^{ij} and summation over i, j leads to

$$(n-1)E_{:k} = -nB_{ik}F_{:j}G^{ij} \tag{12.3}$$

after account is taken of (2.19). It follows from (12.3) and (11.3) that V_n *has constant deviation if and only if either* (1) $F_{:i} = 0$ or (2) $F_{:i} \neq 0$ *and any subspace,* $F =$ *constant, contains each line of curvature with which it has* *a point in common and whose corresponding principal conformal normal curvature differs from zero.* The second alternative implies that the rank of $\| B_{ij} \| < n$ so that at least one of the principal conformal normal curvatures vanishes throughout V_n. This case cannot be realized if $n = 2$ as follows from (3.2).

Conformal differentiation of (12.3) with respect to x^h gives

$$(12.4) \qquad (n-1)E_{:kh} = -n[B_{ik:h}F^i + B_{ik}F_{:jh}G^{ij}].$$

Since $E_{:kh} - E_{:hk} = 0$, (12.4) in conjunction with (2.21) leads to

$$(12.5) \qquad [B_{ik}F_{:jh} - B_{ih}F_{:jk}]G^{ij} = 0.$$

Equations (12.5) are exactly analogous to (11.2) so that a theorem like Theorem 11.1 exists in this case with $F_{:ij}$ occupying the rôle that E_{ij} plays in that theorem.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

BIBLIOGRAPHY.

1. L. P. Eisenhart, *Riemannian Geometry*, 1926, Princeton.
2. A. Fialkow, "Conformal theory of curves," *Proceedings of the National Academy of Sciences*, vol. 26, (1940), pp. 437-439.
3. ———, "Conformal theory of curves," *Transactions of the American Mathematical Society*, vol. 51 (1942), pp. 435-501.
4. ———, "Conformal differential geometry of a subspace," *Transactions of the American Mathematical Society*, vol. 56 (1944), pp. 309-433.

CALCUTTA CENTRAL LIBRARY. UNIVERSITY